

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:47:23 ; Search time 3952.5 Seconds
(without alignments)
17679.337 Million cell updates/sec

Title: US-09-989-981A-5
Perfect score: 2340
Sequence: 1 gtcaggtggagcaggcaggg.....aatattcataaacctatggg 2340

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		%	Query						
No.	Score	Match	Length	DB	ID	Description			
	1	544	23.2	594	9	AV689089	AV689089	AV689089	
	2	521	22.3	597	9	AV694671	AV694671	AV694671	
c	3	477	20.4	477	9	AV720911	AV720911	AV720911	
	4	463.2	19.8	713	10	BB598373	BB598373	BB598373	
c	5	429	18.3	432	9	AI033358	AI033358	ox02f10.s	
c	6	418	17.9	418	9	AI140253	AI140253	qe21a04.x	
	7	407.4	17.4	936	10	BF162656	BF162656	601769307	
	8	396.4	16.9	658	13	BY742680	BY742680	BY742680	
	9	394.4	16.9	417	9	AV695922	AV695922	AV695922	
	10	365	15.6	471	12	BM856449	BM856449	K-EST0140	
	11	360	15.4	360	9	AV660973	AV660973	AV660973	
	12	314.4	13.4	794	14	CA316999	CA316999	UI-M-FW0-	
	13	309.8	13.2	393	9	AI597378	AI597378	vj29c06.y	
	14	307.2	13.1	424	9	AA656720	AA656720	vp95e08.r	
c	15	297	12.7	336	14	T93792	T93792	ye05f01.s1	
	16	288.2	12.3	764	12	BI246567	BI246567	602958477	
	17	286	12.2	722	10	BB667343	BB667343	BB667343	
	18	281.4	12.0	374	9	AA511669	AA511669	vj29c06.r	
	19	281	12.0	460	9	AA239884	AA239884	mx81d01.r	
	20	266.4	11.4	356	14	T86384	T86384	yd77b08.r1	
	21	234.6	10.0	497	14	CD740302	CD740302	4029330 1	
c	22	233.2	10.0	331	14	T86285	T86285	yd77b08.s1	
c	23	211	9.0	625	9	AW112016	AW112016	MC8117 mo	
	24	207	8.8	339	10	BB869579	BB869579	BB869579	
	25	206	8.8	502	9	AA237916	AA237916	mx14e08.r	
	26	206	8.8	535	9	AA244605	AA244605	mx02d10.r	
c	27	203.6	8.7	706	29	AG094162	AG094162	Pan trogl	
	28	199.2	8.5	3623	11	AK004871	AK004871	Mus muscu	
	29	195	8.3	275	9	AI592875	AI592875	vp95e08.y	
	30	188.4	8.1	606	14	CD502116	CD502116	CDA54-H04	
	31	187	8.0	783	13	BX092855	BX092855	BX092855	
	32	179.4	7.7	516	9	AA237183	AA237183	mx17f01.r	
	33	178.6	7.6	366	14	T93842	T93842	ye05f01.r1	
	34	172.4	7.4	2417	11	AK050938	AK050938	Mus muscu	
c	35	161	6.9	1245	14	CD502117	CD502117	CDA54-H04	
c	36	149.8	6.4	837	29	CC710016	CC710016	OGVBI31TH	
	37	147.4	6.3	357	14	CB813243	CB813243	AMGNNUC:T	
	38	147.2	6.3	358	14	CB812866	CB812866	AMGNNUC:T	
c	39	143.8	6.1	861	29	CG262933	CG262933	OG1DH53TV	
	40	142.2	6.1	442	9	AA217272	AA217272	mu89f07.r	
	41	141.6	6.1	695	29	AG122753	AG122753	Pan trogl	
	42	139.8	6.0	833	10	BF620684	BF620684	HVSMEc002	
	43	139	5.9	839	29	CG262656	CG262656	OG1AN46TH	
c	44	135.4	5.8	610	12	BM725726	BM725726	UI-E-EJ0-	
	45	135.4	5.8	749	13	BU740584	BU740584	UI-E-EJ0-	

ALIGNMENTS

RESULT 1

AV689089

LOCUS AV689089 594 bp mRNA linear EST 16-JAN-2002

DEFINITION AV689089 GKC Homo sapiens cDNA clone GKCDZB07 5', mRNA sequence.

ACCESSION AV689089

VERSION AV689089.1 GI:10290952

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 594)

AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

PUBMED 11752456

COMMENT Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES Location/Qualifiers

source

1. .594

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="GKCDZB07"

/tissue_type="hepatocellular carcinoma"

/dev_stage="Adult"

/lab_host="SOLR"

/clone_lib="GKC"

/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN

Query Match 23.2%; Score 544; DB 9; Length 594;

Best Local Similarity 98.9%; Pred. No. 2.7e-118;

Matches 558; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

Qy 1246 GAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCCT 1305

|||||

Db 1 GAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCCT 60

Qy 1306 CCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGT 1365
 |||
 Db 61 CCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGT 120

Qy 1366 AGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAA 1425
 |||
 Db 121 AGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAA 180

Qy 1426 TCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAA 1485
 |||
 Db 181 TCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAA 240

Qy 1486 GTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCAT 1545
 |||
 Db 241 GTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCAT 300

Qy 1546 GATTTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGG 1605
 |||
 Db 301 GATTTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGG 360

Qy 1606 ATATTTTCTGCTGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCT 1665
 |||
 Db 361 ATATTTTCTGCTGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCT 420

Qy 1666 ACTTGGTATCGTCCAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGC 1725
 |||
 Db 421 ACTTGGTATCGTCCAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGC 480

Qy 1726 GGGGGTGCTTGTGGATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAA 1785
 ||
 Db 481 GGNGGTGCTTGTGGATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTT-- 538

Qy 1786 AATCATCAGTTATTTTACATTCCA 1809
 |||
 Db 539 --TCATCAGTTATTNTACATTCCA 560

RESULT 2

AV694671

LOCUS AV694671 597 bp mRNA linear EST 16-JAN-2002

DEFINITION AV694671 GKC Homo sapiens cDNA clone GKCDZG05 5', mRNA sequence.

ACCESSION AV694671

VERSION AV694671.1 GI:10296534

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 597)

AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)


```

          |||
Db      421 ACTTGGTATCGTCC----TCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGC 476
          |||
Qy      1726 GGGGGTGCTTGTGGATCTGGATTCCTCAGAAACATACAAGAAATGCCCATTCCTTTTAA 1785
          |||
Db      477 GGNGGTGCTTGTGGATCTGGATTCCTCAGAAACATACAAGAAATGCCCATTCCTTTTAA 536
          |||

```

RESULT 3

AV720911/c

LOCUS AV720911 477 bp mRNA linear EST 16-OCT-2000

DEFINITION AV720911 GLC Homo sapiens cDNA clone GLCETC06 5', mRNA sequence.

ACCESSION AV720911

VERSION AV720911.1 GI:10818063

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 477)

AUTHORS Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.

TITLE Homo sapiens cDNA GLC clones

JOURNAL Unpublished (2000)

COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source 1. .477
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCETC06"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 20.4%; Score 477; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 2.2e-102;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1843 GTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAAT 1902
          |||
Db      477 GTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAAT 418
          |||
Qy      1903 GTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAG 1962
          |||

```

```

      |||
Db      417 GTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAG 358
Qy      1963 ATTCACAATGAACCTTTCTGATTTTGTATTCAATTTATCCAGCTCTTGTTCATCCTAGGAAT 2022
      |||
Db      357 ATTCACAATGAACCTTTCTGATTTTGTATTCAATTTATCCAGCTCTTGTTCATCCTAGGAAT 298
Qy      2023 AGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAA 2082
      |||
Db      297 AGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAA 238
Qy      2083 ATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCAT 2142
      |||
Db      237 ATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCAT 178
Qy      2143 GTATTTCTTTCTTGACAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCT 2202
      |||
Db      177 GTATTTCTTTCTTGACAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCT 118
Qy      2203 CTTGGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAA 2262
      |||
Db      117 CTTGGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAA 58
Qy      2263 CTTGCAGGGACATGTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAA 2319
      |||
Db      57 CTTGCAGGGACATGTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAA 1

```

RESULT 4

BB598373

LOCUS BB598373 713 bp mRNA linear EST 26-OCT-2001

DEFINITION BB598373 RIKEN full-length enriched, adult male liver tumor Mus musculus cDNA clone C730003G04 5', mRNA sequence.

ACCESSION BB598373

VERSION BB598373.2 GI:16450340

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 713)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT On Dec 1, 2000 this sequence version replaced gi:11506974.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* . 10 (10), 1617-1630 (2000)

wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* . 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* . 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES

source

Location/Qualifiers

1. .713

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="C730003G04"

/sex="male"

/tissue_type="liver tumor"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, adult male liver tumor"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA

was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge."

ORIGIN

Query Match 19.8%; Score 463.2; DB 10; Length 713;
Best Local Similarity 79.3%; Pred. No. 5.1e-99;
Matches 562; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

Qy	90	TTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCATGGGTC	149
Db	2	TTTGCTTCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGC	61
Qy	150	TCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGC	209
Db	62	CTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTCGGTCACGGGCACAGAGG	121
Qy	210	CT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCT	266
Db	122	CTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTCGGGCCTT	181
Qy	267	GGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCT	326
Db	182	GGTGGAAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCT	241
Qy	327	TGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCA	386
Db	242	TGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCA	301
Qy	387	CGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGACCTTCCTGGGGGAGGTGT	446
Db	302	CGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGACCTGGAAGGGGAGGTGT	361
Qy	447	ATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGC	506
Db	362	TTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCTGC	421
Qy	507	AGAGCGACACCCTGCTGAGCAGCCTCACCGTGC GCGAGACGCTGCACTACACCGCGCTGC	566
Db	422	AGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGGAGACGTTGCGATACACAGCGATGC	481
Qy	567	TGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAG	626
Db	482	TGGCCCTCTGCCGAGCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAG	541
Qy	627	AGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATT	686
Db	542	AGCTGAGCCTGAGCCACGTGGCGGACCAATGATTGGCAGCTATAATTTTGGGGGAATTT	601
Qy	687	CCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCA	746
Db	602	CCAGTGGCGAGCGGCGCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCCAGGTCA	661
Qy	747	TGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCA	795

Db 662 TGATGCTAGATGAGCCAACACAGGACTGGACTGCATGNACTGCAATCA 710

RESULT 5

AI033358/c

LOCUS AI033358 432 bp mRNA linear EST 25-JUN-1998

DEFINITION ox02f10.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
clone IMAGE:1655179 3', mRNA sequence.

ACCESSION AI033358

VERSION AI033358.1 GI:3254311

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 432)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 364.

FEATURES Location/Qualifiers

source

1. .432

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1655179"

/sex="male"

/dev_stage="20 week-post conception fetus"

/lab_host="DH10B (ampicillin resistant)"

/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'

AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 18.3%; Score 429; DB 9; Length 432;

Best Local Similarity 100.0%; Pred. No. 5.7e-91;

Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1908 CCTTCACTCAAGGAATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCA 1967

|||||

Db 432 CCTTCACTCAAGGAATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCA 373

Qy 1968 CAATGAAC TTTCTGATTTTGTATTTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTG 2027
 |||
 Db 372 CAATGAAC TTTCTGATTTTGTATTTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTG 313

Qy 2028 TTTTCAAAATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGA 2087
 |||
 Db 312 TTTTCAAAATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGA 253

Qy 2088 AGTGAAGCTGCCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATT 2147
 |||
 Db 252 AGTGAAGCTGCCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATT 193

Qy 2148 TCTTTCTTGACAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGG 2207
 |||
 Db 192 TCTTTCTTGACAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGG 133

Qy 2208 ATCCAAGCAGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTTGC 2267
 |||
 Db 132 ATCCAAGCAGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTTGC 73

Qy 2268 AGGGACATGTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTC 2327
 |||
 Db 72 AGGGACATGTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTC 13

Qy 2328 ATAAACCTA 2336
 |||
 Db 12 ATAAACCTA 4

RESULT 6

AI140253/c

LOCUS AI140253 418 bp mRNA linear EST 29-OCT-1998

DEFINITION qe2la04.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
 IMAGE:1739598 3', mRNA sequence.

ACCESSION AI140253

VERSION AI140253.1 GI:3647710

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 418)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1828 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 417.

FEATURES Location/Qualifiers

source 1..418

/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1739598"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_fetal_lung_NbHL19W"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHH19W."

```

ORIGIN

```

Query Match          17.9%; Score 418; DB 9; Length 418;
Best Local Similarity 100.0%; Pred. No. 2.3e-88;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1918 AGGAATTC AATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCACAATGAACTT 1977
          |||
Db      418 AGGAATTC AATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCACAATGAACTT 359

Qy      1978 TCTGATTTTGTATTCAATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAAAT 2037
          |||
Db      358 TCTGATTTTGTATTCAATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAAAT 299

Qy      2038 AAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTG 2097
          |||
Db      298 AAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTG 239

Qy      2098 CCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGA 2157
          |||
Db      238 CCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGA 179

Qy      2158 CAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAG 2217
          |||
Db      178 CAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAG 119

Qy      2218 GCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTTGCAGGGACATGT 2277
          |||
Db      118 GCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTTGCAGGGACATGT 59

Qy      2278 GGT TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCT 2335
          |||
Db      58 GGT TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCT 1

```

RESULT 7

BF162656

```

LOCUS      BF162656          936 bp    mRNA    linear    EST 30-OCT-2000
DEFINITION  601769307F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3988777 5',

```


mRNA sequence.

ACCESSION BF162656

VERSION BF162656.1 GI:11042879

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 936)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9197 row: m column: 02
High quality sequence stop: 686.

FEATURES Location/Qualifiers

source 1. .936
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:3988777"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu29"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

ORIGIN

Query Match 17.4%; Score 407.4; DB 10; Length 936;
Best Local Similarity 75.2%; Pred. No. 1.1e-85;
Matches 534; Conservative 0; Mismatches 172; Indels 4; Gaps 2;

```

Qy      605 AAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGC 664
          | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      127 AGGGTAGAGGCAGTCATGACAGAGCTGAGTCTGAGCCACGTGGCGGACCAAATGATTGGC 186

Qy      665 AACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAG 724
          | ||| | || ||||| ||||| || ||||| ||||| || ||||| |||||
Db      187 AGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGCAGCCCAA 246

Qy      725 CTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATG 784
          || || ||||| || ||||| || || ||||| ||||| ||||| |||||
Db      247 CTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATG 306

Qy      785 ACTGCTAATCAGATTGTTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTT 844

```

Db 307 ACTGCAAATCAAATGGTCCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATT 366
 Qy 845 CTCACCATTCACCAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCATCCTG 904
 Db 367 GTCACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTG 426
 Qy 905 AGCTTCGGAGAGCTGATTTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAATGAC 964
 Db 427 ACTTACGGAGAGTNGGTGTTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTTCAATAAC 486
 Qy 965 TCGGGTTACCCCTGTCTGAACATTCAAACCCCTTTGACTTCTATATGGACCTGACGTCA 1024
 Db 487 TGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTGATTTTACATGGACTTGACATCA 546
 Qy 1025 GTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAA 1084
 Db 547 GTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTACAAGCGAGTACAGATGCTGGAC 606
 Qy 1085 TCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAG--AATATTGAAAGAATGA 1142
 Db 607 TGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTGGAGACACATTGCACAGAGCAC 666
 Qy 1143 AACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTTCT 1202
 Db 667 GATACCTGAAAACCTTAACCACGGTTCCTTTCAAACAAAAGATCTCCTGGGATGTTTCG 726
 Qy 1203 CTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACT--TGGTGAGAAATAAGCTGGC 1260
 Db 727 GCCAGCTTGGGGTCCTGGAGAGGGAATTACAAGAAACCTCCACGCGGAATAAGCACGGC 786
 Qy 1261 AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTT 1310
 Db 787 ACGGATAAATGCGCCACGGCAGAACTCGGTCACGGGCCTTACCACATAT 836

RESULT 8

BY742680

LOCUS BY742680 658 bp mRNA linear EST 17-DEC-2002

DEFINITION BY742680 RIKEN full-length enriched, adult male liver tumor Mus musculus cDNA clone C730040P06 5', mRNA sequence.

ACCESSION BY742680

VERSION BY742680.1 GI:27168376

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 658)

AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,

Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasaki,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perteau,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
source Location/Qualifiers
1. .658
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="C730040P06"
/sex="male"
/tissue_type="liver tumor"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male liver tumor"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge."

ORIGIN

Query Match 16.9%; Score 396.4; DB 13; Length 658;
Best Local Similarity 78.2%; Pred. No. 4e-83;
Matches 513; Conservative 0; Mismatches 138; Indels 5; Gaps 3;

```
Qy      90 TTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCATGGGGTC 149
      ||||| |||| | ||||| |||| | | | || | | | || | || |
Db      2 TTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGC 61

Qy     150 TCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGC 209
      || | ||||| || | ||||| ||||| | | | || | | | || |
Db     62 CTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGG 121

Qy     210 CT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCT 266
      || ||||| | || |||| |||| ||||| ||||| |||| | |||| |
Db    122 CTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTCGGGCCTT 181

Qy     267 GGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCT 326
      |||| | |||| || | |||| |||| | |||| | ||||| ||||| |||||
```

Db 182 GGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCT 241

Qy 327 TGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCA 386
 ||||| | ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 242 TGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCA 301

Qy 387 CGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGT 446
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 302 CGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGT 361

Qy 447 ATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCTGC 506
 ||||| ||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 362 TTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCTGC 421

Qy 507 AGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGC 566
 ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 422 AGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGCGATGC 481

Qy 567 TGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAG 626
 ||||| || ||||| || | | || || | | ||||| ||||| ||||| |||||

Db 482 TGGCCCTCTGCCGCAGCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAG 541

Qy 627 AGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTT 686
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 542 AGCTGAGCCTGAGCCACGTGGCGGACCANATGATTGGCAGCTATAAATTTGGGGG-ATNT 600

Qy 687 CCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAG 742
 ||| || ||||| ||||| || ||||| ||||| || || | || || |

Db 601 CCAGTGGCGAGCGGCGCCGAGT-TCCATCGCAGCCCAACTCTTCAGGACCCCAAGG 655

RESULT 9

AV695922

LOCUS AV695922 417 bp mRNA linear EST 16-JAN-2002

DEFINITION AV695922 GKC Homo sapiens cDNA clone GKCDWE04 5', mRNA sequence.

ACCESSION AV695922

VERSION AV695922.1 GI:10297785

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 417)

AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

PUBMED 11752456

COMMENT Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

ORIGIN

RESULT 10

```

LOCUS       BM856449                471 bp    mRNA    linear    EST 06-MAR-2002
DEFINITION  K-EST0140406 S14K402 Homo sapiens cDNA clone S14K402-48-E04 5',
            mRNA sequence.
ACCESSION   BM856449

```

VERSION BM856449.1 GI:19212848
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 471)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 48 row: E column: 04
 High quality sequence stop: 471.

FEATURES
 source Location/Qualifiers
 1. .471
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S14K402-48-E04"
 /cell_line="K402"
 /lab_host="Top10F"
 /clone_lib="S14K402"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
 Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dT-selected mRNA by
 priming with dT-tailed vector. The dT-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F' by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

ORIGIN

Query Match 15.6%; Score 365; DB 12; Length 471;
 Best Local Similarity 100.0%; Pred. No. 1e-75;
 Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1976 TTTCTGATTTTGTATTCATTTATTCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAA 2035
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 TTTCTGATTTTGTATTCATTTATTCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAA 60
 Qy 2036 ATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGC 2095

```

      |||
Db      61 ATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGC 120
Qy      2096 TGCCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTT 2155
      |||
Db      121 TGCCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTT 180
Qy      2156 GACAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGC 2215
      |||
Db      181 GACAGGACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGC 240
Qy      2216 AGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACAT 2275
      |||
Db      241 AGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACAT 300
Qy      2276 GTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCT 2335
      |||
Db      301 GTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCT 360
Qy      2336 ATGGG 2340
      ||||
Db      361 ATGGG 365

```

RESULT 11

AV660973

LOCUS AV660973 360 bp mRNA linear EST 16-JAN-2002

DEFINITION AV660973 GLC Homo sapiens cDNA clone GLCGNC08 3', mRNA sequence.

ACCESSION AV660973

VERSION AV660973.1 GI:9881987

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 360)

AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

PUBMED 11752456

COMMENT Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES Location/Qualifiers

source 1. .360

/organism="Homo sapiens"


```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCGNC08"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

```

ORIGIN

```

Query Match          15.4%;  Score 360;  DB 9;  Length 360;
Best Local Similarity 100.0%;  Pred. No. 1.4e-74;
Matches 360;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1577 GGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCCAC 1636
          |||
Db      1    GGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCCAC 60

Qy      1637 TTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTC 1696
          |||
Db      61    TTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTC 120

Qy      1697 AACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCTCCTCAGA 1756
          |||
Db      121   AACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCTCCTCAGA 180

Qy      1757 AACATACAAGAAATGCCCATTCTTTTAAAATCATCAGTTATTTTACATTCCAAAAATAT 1816
          |||
Db      181   AACATACAAGAAATGCCCATTCTTTTAAAATCATCAGTTATTTTACATTCCAAAAATAT 240

Qy      1817 TGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCA 1876
          |||
Db      241   TGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCA 300

Qy      1877 AATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAG 1936
          |||
Db      301   AATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAG 360

```

RESULT 12

CA316999

LOCUS CA316999 794 bp mRNA linear EST 09-JUL-2003

DEFINITION UI-M-FW0-cbm-a-08-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
IMAGE:6811377 5', mRNA sequence.

ACCESSION CA316999

VERSION CA316999.1 GI:24535123

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 794)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

FEATURES	Location/Qualifiers
source	1. .794 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:6811377" /tissue_type="whole brain" /dev_stage="embryo 13.5,14.5,16.5,17.5dpc" /lab_host="DH10B (T1 phage resistant)" /clone_lib="NIH_BMAP_FW0" /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 13.4%; Score 314.4; DB 14; Length 794;
 Best Local Similarity 71.8%; Pred. No. 1.5e-63;
 Matches 481; Conservative 0; Mismatches 103; Indels 86; Gaps 2;

Qy	370	AGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGAC	429
Db	87	AGGCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGAC	146
Qy	430	CTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTG	489
Db	147	CCTGGAAGGGGAGGTGTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTG	206
Qy	490	CTTCTCCTACG-----	500
Db	207	CTTCTCCTACGTCCTGCAGGTGGGCGTGTCCTGGCCCTAGCCTGCCCGGGCTCTGGCCC	266
Qy	501	-----TCCTGCAGAGCGACACCCTGCTGA	524

Db	267	CTAGCCCCGGGATTTTCGACGACCCCTGATGTCCCTTTCTGCAGAGCGACGTTTTTCTGA	326
Qy	525	GCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCA	584
Db	327	GCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGCGAGCT	386
Qy	585	ATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATG	644
Db	387	CCGCGGACTTCTACAACAAGAAGGTANAGGCAGTCATGACAGAGCTGAGCCTGAGCCACG	446
Qy	645	TGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCC	704
Db	447	TGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGCC	506
Qy	705	GGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAA	764
Db	507	GAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAA	566
Qy	765	CCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTC	824
Db	567	CCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCTC	626
Qy	825	GCAGGAACCGAATTGTGGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCT	884
Db	627	GCAGGGACCGAATTGTGATTGTCACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACT	686
Qy	885	TTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAA	944
Db	687	TCGACANAATTGCCATCCTGACTTACGGAGAGTTGGTGTCTGTGGCAC-CCAGAGGAGA	745
Qy	945	TGCTTGATTT 954	
Db	746	TGCTGGTTCT 755	

AI597378

LOCUS AT597378 393 bp mRNA linear EST 21-APR-1999

DEFINITION vj29c06.y1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA clone IMAGE:930442 5', mRNA sequence.

ACCESSION AT597378

VERSION AI597378.1 GI:4606426

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

SOURCE	Has mascatus
ORGANISM	Mus musculus

REFERENCE 1 (bases 1 to 393)

AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE	The WashU-NCI Mouse EST Project 1999
-------	--------------------------------------

JOURNAL Unpublished (1999)

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:535362

This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)

Seq primer: -40RP from Gibco

High quality sequence stop: 389.

FEATURES	Location/Qualifiers
source	1. .393
	/organism="Mus musculus"
	/mol_type="mRNA"
	/db_xref="taxon:10090"
	/clone="IMAGE:930442"
	/tissue_type="diaphragm"
	/dev_stage="adult"
	/lab_host="SOLR (kanamycin resistant)"
	/clone_lib="Stratagene mouse diaphragm (#937303)"
	/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dT. Average insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match	13.2%;	Score 309.8;	DB 9;	Length 393;
Best Local Similarity	86.8%;	Pred. No. 1.4e-62;		
Matches 341;	Conservative 0;	Mismatches 52;	Indels 0;	Gaps 0;

Qy	1425	ATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGA	1484
Db	1	ATCTGTTTCCCATGCTGAGAGCCGTCAGCGACCAGGAGAGTCAGGATGGCCTGTATCATA	60
Qy	1485	AGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCCTCCCTTCAGCGTTGTTGCCACCA	1544
Db	61	AGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCCTCCCTTCAGCGTCATCGCCACGG	120
Qy	1545	TGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTTG	1604
Db	121	TCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTTGATCCTGAAGTTGCCAGATTTG	180
Qy	1605	GATATTTTCTGCTGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGC	1664
Db	181	GATATTTCTGCTGCTCTTTTGGCCCCCACTTAATTGGAGAATTTCTAACACTTGTGC	240
Qy	1665	TACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTG	1724
Db	241	TGCTTGGTATAGTCCAAAACCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATCT	300
Qy	1725	CGGGGGTGCTTGTGGATCTGGATTCCTCAGAAACATACAAGAAATGCCCATTCCTTTTA	1784
Db	301	CTGGGCTGCTTATTGGATCTGGATTTATCAGAAACATACAAGAAATGCCCATTCCTTTAA	360

Qy 1785 AAATCATCAGTTATTTTACATTCCAAAAATATT 1817
 ||||| | |||||
 Db 361 AAATCCTGGGTTATTTTACATTCCAAAAATACT 393

RESULT 14

AA656720

LOCUS AA656720 424 bp mRNA linear EST 04-NOV-1997

DEFINITION vp95e08.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA clone IMAGE:1092518 5' similar to SW:SCRT_DROME P45843 SCARLET PROTEIN. ;, mRNA sequence.

ACCESSION AA656720

VERSION AA656720.1 GI:2592874

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 424)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:598750

Seq primer: -28m13 rev1 ET from Amersham.

FEATURES Location/Qualifiers

source

1. .424

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:1092518"

/tissue_type="diaphragm"

/dev_stage="adult"

/lab_host="SOLR (kanamycin resistant)"

/clone_lib="Stratagene mouse diaphragm (#937303)"

/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA

prepared from diaphragm muscle. Primer: Oligo dT. Average

insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor

sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match

13.1%; Score 307.2; DB 9; Length 424;

Best Local Similarity 82.8%; Pred. No. 5.8e-62;
Matches 351; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

Qy      1186 TTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGT 1245
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1    TCCTCCTGGGATGTTCTGGCAAGCTTGGTGTCTGCTGAGGCGAGTAACAAGAACTTAAT 60

Qy      1246 GAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCT 1305
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61    GAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCAGAATCTGATCATGGGCCTCTTCT 120

Qy      1306 CCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGT 1365
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      121   CATTTTCTACCTTCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGCGT 180

Qy      1366 AGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAA 1425
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      181   GGGGCTGCTCTATCAGCTTGTGGGTGCCACCCCATACACCGGCATGCTCAATGCTGTGAA 240

Qy      1426 TCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAA 1485
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      241   TCTGTTTCCCATGCTGAGAGCCGTGAGCGACCAGGAGAGTCAGGATGGCCTGTATCATAA 300

Qy      1486 GTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCTTCAGCGTTGTTGCCACCAT 1545
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      301   GTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCTCCCTTCAGCGTCATCGCCACGGT 360

Qy      1546 GATTTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCCGATTTGG 1605
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      361   CATTTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTTGTATCCTGAAGTTGCCAGATTTGG 420

Qy      1606 ATAT 1609
          | | | |
Db      421   ATAT 424

```

RESULT 15

T93792/c

LOCUS T93792 336 bp mRNA linear EST 23-MAR-1995

DEFINITION ye05f01.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone
IMAGE:116857 3', mRNA sequence.

ACCESSION T93792

VERSION T93792.1 GI:726965

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 336)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK

High quality sequence stops: 265 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 768 Std Error: 0.00
Seq primer: -21ml3
High quality sequence stop: 265.

ORIGIN

Qy	2014	CCTAGGAATAGTTGTTTTCAAATAAAGGG--ATCATCTCATTAGCAGGTAGTGAAAGCCA	2071
Db	322	CCTAGGAANAGTTGNNTTCAAATAAAGGGGATCATCCTCA TTAGCAGGTAGTGAAAGCCA	263
Qy	2072	TGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGAACGTCTGAAAT	2131
Db	262	TGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGAACGTCTGAAAT	203
Qy	2132	GAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTCAAGTC TTTTAACCATT AAGACTC	2191
Db	202	GAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTCAAGTC TTTTAACCATT AAGACTC	143
Qy	2192	CATTTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCT	2251
Db	142	CATTTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGTTTATAGTCCCT	83
Qy	2252	TGCTCTTACA ACTTG CAGGGACATGTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAG	2311
Db	82	TGCTCTTACA ACTTG CAGGGACATGTGGTTATTTGGAAATTGTGACTGAGCGGACCCAAG	23

```
Qy      2312 AATGTAAATAATATTCATAAAC 2333
          |||
Db      22  AATGTAAATAATATTCATAANC 1
```

Search completed: February 26, 2004, 09:39:30
Job time : 3959.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:40:23 ; Search time 6010.48 Seconds
(without alignments)
16874.299 Million cell updates/sec

Title: US-09-989-981A-5
Perfect score: 2340
Sequence: 1 gtcaggtggagcaggcaggg.....aatattcataaacctatggg 2340

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*

```

28:  em_un:*
29:  em_vi:*
30:  em_htg_hum:*
31:  em_htg_inv:*
32:  em_htg_other:*
33:  em_htg_mus:*
34:  em_htg_pln:*
35:  em_htg_rod:*
36:  em_htg_mam:*
37:  em_htg_vrt:*
38:  em_sy:*
39:  em_htgo_hum:*
40:  em_htgo_mus:*
41:  em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			%		Query		DB	ID	Description
	No.	Score	Match	Length	Match	Length			
	1	2340	100.0	2340	6	AX320883		AX320883	Sequence
	2	2340	100.0	2340	6	AX685733		AX685733	Sequence
	3	2340	100.0	2340	9	AF320293		AF320293	Homo sapi
	4	2340	100.0	2740	9	AF312715		AF312715	Homo sapi
	5	2338.4	99.9	2516	6	AX456520		AX456520	Sequence
	6	1920	82.1	1920	6	AX456519		AX456519	Sequence
	7	1410.8	60.3	2351	10	AY195873		AY195873	Mus muscu
	8	1409.2	60.2	2354	6	AX456524		AX456524	Sequence
	9	1409.2	60.2	2354	10	AF312713		AF312713	Mus muscu
	10	1406	60.1	2351	10	AY195872		AY195872	Mus muscu
	11	1395.6	59.6	2258	6	AX320881		AX320881	Sequence
	12	1383.8	59.1	2470	10	AF312714		AF312714	Rattus no
	13	1365.4	58.4	1959	6	AX685729		AX685729	Sequence
	14	1363	58.2	2035	6	AX456526		AX456526	Sequence
	15	1335.8	57.1	1915	6	AX456523		AX456523	Sequence
	16	1174.2	50.2	2512	6	AX747300		AX747300	Sequence
	17	1174.2	50.2	2512	9	AK091997		AK091997	Homo sapi
	18	770.2	32.9	1069	6	AX456528		AX456528	Sequence
	19	473	20.2	186091	9	AC011242		AC011242	Homo sapi
	20	472	20.2	472	6	AX320898		AX320898	Sequence
c	21	357.2	15.3	178016	2	AC146787		AC146787	Aotus nan
c	22	355.6	15.2	185045	2	AC146466		AC146466	Callithri
c	23	352.4	15.1	207760	2	AC146286		AC146286	Callicebu
	24	341.8	14.6	202533	2	AC146464		AC146464	Saimiri s
c	25	312	13.3	127066	9	AC084265		AC084265	Homo sapi
c	26	312	13.3	139342	9	AC108476		AC108476	Homo sapi
	27	271.6	11.6	159346	2	AC145533		AC145533	Lemur cat
	28	268	11.5	281	6	BD234212		BD234212	ATP-bindi
c	29	249.6	10.7	2809	9	F351812S01		AF351812	Homo sapi
	30	249.6	10.7	4899	9	AF404106		AF404106	Homo sapi
	31	249.6	10.7	5459	6	AX456521		AX456521	Sequence
	32	249	10.6	249	6	AX320886		AX320886	Sequence
c	33	249	10.6	581	9	AF404107		AF404107	Homo sapi

34	230.2	9.8	68166	2	AC084712	AC084712 Homo sapi
35	214	9.1	214	6	AX320893	AX320893 Sequence
36	206	8.8	206	6	AX320894	AX320894 Sequence
37	203.6	8.7	2022	9	AF320294	AF320294 Homo sapi
38	203.6	8.7	2669	6	AX685735	AX685735 Sequence
39	203.6	8.7	2679	9	AF324494	AF324494 Homo sapi
40	200.8	8.6	2285	10	AY196215	AY196215 Mus muscu
41	199.2	8.5	2019	6	AX685731	AX685731 Sequence
42	199.2	8.5	2284	10	AY196216	AY196216 Mus muscu
43	199.2	8.5	3674	10	AF324495	AF324495 Mus muscu
44	189.2	8.1	4829	10	AF351785	AF351785 Rattus no
45	186	7.9	186	6	AX320896	AX320896 Sequence

ALIGNMENTS

RESULT 1

AX320883

LOCUS AX320883 2340 bp DNA linear PAT 14-DEC-2001

DEFINITION Sequence 4 from Patent WO0179272.

ACCESSION AX320883

VERSION AX320883.1 GI:17902433

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Tian,H., Schultz,J. and Shan,B.

TITLE Sitosterolemia susceptibility gene (ssg): compositions and methods of use

JOURNAL Patent: WO 0179272-A 4 25-OCT-2001; Tularik Inc. (US)

FEATURES Location/Qualifiers

source 1. .2340

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

/note="human sitosterolemia gene (SSG)"

CDS 107. .2062

/note="unnamed protein product; human sitosterolemia susceptibility gene (SSG) protein"

/codon_start=1

/protein_id="CAD19409.1"

/db_xref="GI:17902434"

/db_xref="REMTREMBL:CAD19409"

/translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
YSVSHRVRPWWDDITSCRQQWTRQILKDVSLYVESGQIMCILGSSGSKTTLDDAMSGR
LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRG
PGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIQAQLLQDPKVMFLFDE
PTTGLDCMTANQIVVLLVELARRNRIVVLTIIHQPRSELFQLFDKIALLSFGELIFCGT
PAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS
CHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRLRVTRNLVRNKLAVITRLQ
NLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMNAVNLFPVLRVSD
QESQDGLYQKWQMLAYALHVLFPFSVVATMIFSSVCYWTGLGHPEVARFGYFSAALLA
PHLIGEFLLTLVLLGIVQNPNIIVNSVVALLSIAGVLVGSGLRNIQEMPIPFKIISYFT

FQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF
LILYSFIPALVILGIVVFKIRDHLISR"

ORIGIN

Query Match 100.0%; Score 2340; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	60
Db	1	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	60
Qy	61	GGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Db	61	GGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Qy	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCT	180
Db	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCT	180
Qy	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Db	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Qy	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTACCGTGCG	540
Db	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTACCGTGCG	540
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Qy	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCGGGTCTCCATCGCAGC	720
Db	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCGGGTCTCCATCGCAGC	720
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780

Db	721	 CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCTGGACTG	780
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACCTGGCTCGCAGGAACCGAATTGT	840
Db	781	 CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACCTGGCTCGCAGGAACCGAATTGT	840
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	841	 GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGAAATGCTTGATTTCTTCAA	960
Db	901	 CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGAAATGCTTGATTTCTTCAA	960
Qy	961	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Db	961	 TGACTGCGGTTACCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1021	 GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1081	 AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1141	 GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1201	 CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Db	1261	 AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Db	1321	 GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1381	 GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1441	 GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Qy	1501	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1501	 GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTGGATATTTTCTGCTGC	1620

```

Db      1561 GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCCGATTGGATATTTTCTGCTGC 1620
Qy      1621 TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA 1680
          |||
Db      1621 TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA 1680
Qy      1681 AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG 1740
          |||
Db      1681 AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG 1740
Qy      1741 ATCTGGATTCCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT 1800
          |||
Db      1741 ATCTGGATTCCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT 1800
Qy      1801 TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT 1860
          |||
Db      1801 TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT 1860
Qy      1861 CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG 1920
          |||
Db      1861 CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG 1920
Qy      1921 AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACCTTCT 1980
          |||
Db      1921 AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACCTTCT 1980
Qy      1981 GATTTTGTATTCAATTTATTCCAGCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAG 2040
          |||
Db      1981 GATTTTGTATTCAATTTATTCCAGCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAG 2040
Qy      2041 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2100
          |||
Db      2041 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2100
Qy      2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160
          |||
Db      2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160
Qy      2161 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
          |||
Db      2161 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
Qy      2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
          |||
Db      2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
Qy      2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
          |||
Db      2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340

```

RESULT 2

AX685733

LOCUS AX685733 2340 bp DNA linear PAT 29-MAR-2003

DEFINITION Sequence 5 from Patent WO02081691.

ACCESSION AX685733

VERSION AX685733.1 GI:29371742

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.

TITLE Abcg5 and abcg8: compositions and methods of use

JOURNAL Patent: WO 02081691-A 5 17-OCT-2002;
Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)

FEATURES Location/Qualifiers

source 1. .2340
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

CDS 107. .2062
/note="unnamed protein product; human ABCG5 (hABCG5)"
/codon_start=1
/protein_id="CAD86572.1"
/db_xref="GI:29371743"
/db_xref="REMTREMBL:CAD86572"
/translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
YSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLDDAMSGR
LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGN
PGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVISAAQLLQDPKVMFLDE
PTTGLDCMTANQIVVLLVELARRNRIVLTIHQPRSELFQLFDKIAILSFGEIIFCGT
PAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS
CHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQ
NLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVLRVAVSD
QESQDGLYQKWQMMLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLA
PHLIGEFLLVLLGIVQNPINVSVALLSIAGVLVSGSGLRNIQEMPIPFKIIISYFT
FQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF
LILYSFIPALVILGIVVFKIRDHLISR"

ORIGIN

Query Match 100.0%; Score 2340; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	60
Db	1	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	60
Qy	61	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Db	61	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Qy	121	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Db	121	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Qy	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300

Db	241	 CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	301	 CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	361	 AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Qy	421	CGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	421	 CGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Db	481	 CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	541	 CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Qy	601	GAAGAAGGTGGAGGCCGTGATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	601	 GAAGAAGGTGGAGGCCGTGATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	661	 TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	721	 CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	781	 CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	841	 GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	901	 CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	960
Qy	961	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Db	961	 TGACTGCGGTTACCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1021	 GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140

Db	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTTCCTCCTTTCTTCGTTCT	1320
Db	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTTCCTCCTTTCTTCGTTCT	1320
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Db	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Qy	1501	GGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1501	GGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTCTGCTGC	1620
Db	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTCTGCTGC	1620
Qy	1621	TCTCTTGGCCCCCCTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1680
Db	1621	TCTCTTGGCCCCCCTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1680
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG	1740
Db	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG	1740
Qy	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCATTCTCTTTTAAATCATCAGTTATTT	1800
Db	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCATTCTCTTTTAAATCATCAGTTATTT	1800
Qy	1801	TACATTCCAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1801	TACATTCCAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Qy	1861	CACCTGTGGCAGCTCAAATGTTTCTGTGACAATAATCCAATGTGTGCCTTCACTCAAGG	1920
Db	1861	CACCTGTGGCAGCTCAAATGTTTCTGTGACAATAATCCAATGTGTGCCTTCACTCAAGG	1920
Qy	1921	AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACCTTCT	1980
Db	1921	AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACCTTCT	1980

Qy 1981 GATTTTGTATTTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAATAAG 2040
 |||
 Db 1981 GATTTTGTATTTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAATAAG 2040

Qy 2041 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2100
 |||
 Db 2041 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2100

Qy 2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160
 |||
 Db 2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160

Qy 2161 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
 |||
 Db 2161 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220

Qy 2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
 |||
 Db 2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280

Qy 2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
 |||
 Db 2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340

RESULT 3

AF320293

LOCUS AF320293 2340 bp mRNA linear PRI 13-DEC-2000

DEFINITION Homo sapiens ABCG5 (ABCG5) mRNA, complete cds.

ACCESSION AF320293

VERSION AF320293.1 GI:11692799

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2340)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J.,
 Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Accumulation of Dietary Cholesterol in Sitosterolemia Caused by
 Mutations in Adjacent ABC Transporters

JOURNAL Science (2001) In press

REFERENCE 2 (bases 1 to 2340)

AUTHORS Berge,K.E., Tian,H., Graf,G.A., Yu,L., Grishin,N.V., Schultz,J.,
 Kwiterovich,P., Shan,B., Barnes,R. and Hobbs,H.H.

TITLE Direct Submission

JOURNAL Submitted (09-NOV-2000) Molecular Genetics, University of Texas,
 Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd.,
 Dallas, TX 75390-9046, USA

FEATURES Location/Qualifiers

source 1. .2340

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

gene 1. .2340

/gene="ABCG5"

CDS

```
107. .2062
/gene="ABCG5"
/note="ATP-binding cassette, subfamily G, member 5"
/codon_start=1
/product="ABCG5"
/protein_id="AAG40003.1"
/db_xref="GI:11692800"
/translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
YSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLDDAMSGR
LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLYHTALLAIRRG
PGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLDE
PTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGT
PAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKSAI
CHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQ
NLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVLRVSD
QESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLA
PHLIGEFLLVLLGIVQNPNIIVNSVVALLSIAGVLVSGSGLRNIQEMPIPFKIISYFT
FQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF
LILYSFIPALVILGIVVFKIRDHLISR"
```

ORIGIN

```
Query Match          100.0%;  Score 2340;  DB 9;  Length 2340;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 2340;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
Qy      1  GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60

Qy     61  GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120

Qy    121  ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180

Qy    181  GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240

Qy    241  CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC 300

Qy    301  CAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGGGCAGATCATGTGCATCCT 360
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGGGCAGATCATGTGCATCCT 360

Qy    361  AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG 420
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG 420

Qy    421  CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT 480
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT 480
```

Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Db	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Qy	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Qy	961	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Db	961	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Db	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCTAGGTCTCCTTTACCA	1380

Db	1321	 GCGGGTCCGAAGCAATGTGCTAAAGGGTGTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1381	 GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1441	 GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Qy	1501	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1501	 GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTGGATATTTTCTGCTGC	1620
Db	1561	 GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTGGATATTTTCTGCTGC	1620
Qy	1621	TCTCTTGGCCCCCCTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1680
Db	1621	 TCTCTTGGCCCCCCTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1680
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG	1740
Db	1681	 AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG	1740
Qy	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT	1800
Db	1741	 ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT	1800
Qy	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1801	 TACATTCCAAAAATATTGCAGTGAGATTCTTGTAAGTCAATGAGTTCTACGGACTGAATTT	1860
Qy	1861	CACCTGTGGCAGCTCAAATGTTTCTGTGACAACCTAATCCAATGTGTGCCTTCACTCAAGG	1920
Db	1861	 CACCTGTGGCAGCTCAAATGTTTCTGTGACAACCTAATCCAATGTGTGCCTTCACTCAAGG	1920
Qy	1921	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCACAATGAACCTTTCT	1980
Db	1921	 AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCACAATGAACCTTTCT	1980
Qy	1981	GATTTTGTATTCAATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAAATAAG	2040
Db	1981	 GATTTTGTATTCAATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAAATAAG	2040
Qy	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Db	2041	 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Qy	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160
Db	2101	 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160
Qy	2161	GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC	2220

Db 2161 GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220

Qy 2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
 |||

Db 2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280

Qy 2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
 |||

Db 2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340

RESULT 4

AF312715

LOCUS AF312715 2740 bp mRNA linear PRI 14-JUN-2001

DEFINITION Homo sapiens sterolin (ABCG5) mRNA, complete cds.

ACCESSION AF312715

VERSION AF312715.2 GI:14423628

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2740)

AUTHORS Lee,M.H., Lu,K., Hazard,S., Yu,H., Shulenin,S., Hidaka,H.,
 Kojima,H., Allikmets,R., Sakuma,N., Pegoraro,R., Srivastava,A.K.,
 Salen,G., Dean,M. and Patel,S.B.

TITLE Identification of a gene, ABCG5, important in the regulation of
 dietary cholesterol absorption

JOURNAL Nat. Genet. 27 (1), 79-83 (2001)

MEDLINE 20578753

PUBMED 11138003

REFERENCE 2 (bases 1 to 2740)

AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB541, Charleston, SC 29403, USA

COMMENT On Jun 14, 2001 this sequence version replaced gi:12382303.

FEATURES Location/Qualifiers

source

1. .2740
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2p21; between D2S2294 and D2S2298"
 /tissue_type="liver"

gene

1. .2740
 /gene="ABCG5"

CDS

141. .2096
 /gene="ABCG5"
 /codon_start=1
 /product="sterolin"
 /protein_id="AAG53099.1"
 /db_xref="GI:12382304"
 /translation="MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHAS
 YSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGR
 LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGN"

PGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDE
 PTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGEIIFCGT
 PAEMLDFFNDNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS
 AI
 CHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLQ
 NLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSD
 QESQDGLYQKWQMLLAYALHVLPPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLA
 PHLIGEFLTLVLLGIVQNPNIIVNSVALLSIAGVLVGSGLRNIQEMPIPFKIIISYFT
 FQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF
 LILYSFIPALVILGIVVFKIRDHLISR"

ORIGIN

Query Match	100.0%;	Score 2340;	DB 9;	Length 2740;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2340;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	60
Db	35	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	94
Qy	61	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Db	95	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	154
Qy	121	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Db	155	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	214
Qy	181	GGAGGGGGCTCCTGCCACCGCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	215	GGAGGGGGCTCCTGCCACCGCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	274
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Db	275	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	334
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	335	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	394
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	395	AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	454
Qy	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	455	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	514
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCCTGCG	540
Db	515	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCCTGCG	574
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	575	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	634
Qy	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660

Db	635	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	694
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	695	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	754
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	755	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	814
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	815	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	874
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	875	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCAT	934
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	935	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	994
Qy	961	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1020
Db	995	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1054
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1055	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1114
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1115	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAAACTTTGAAGAATATTGAAAGAAT	1174
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1175	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1234
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1235	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1294
Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Db	1295	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1354
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGGTAGGTCTCCTTTACCA	1380
Db	1355	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGGTAGGTCTCCTTTACCA	1414
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1415	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1474
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1475	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1534

Qy	1501	GGCCTATGCACTGCACGTCCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1535	GGCCTATGCACTGCACGTCCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1594
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGGATATTTTCTGCTGC	1620
Db	1595	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGGATATTTTCTGCTGC	1654
Qy	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1680
Db	1655	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1714
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGTGCTTGTGG	1740
Db	1715	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGTGCTTGTGG	1774
Qy	1741	ATCTGGATTCCCTCAGAAACATACAAGAAATGCCATTCCCTTTTAAAATCATCAGTTATTT	1800
Db	1775	ATCTGGATTCCCTCAGAAACATACAAGAAATGCCATTCCCTTTTAAAATCATCAGTTATTT	1834
Qy	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1835	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1894
Qy	1861	CACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGG	1920
Db	1895	CACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGG	1954
Qy	1921	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACAAATGAACCTTTCT	1980
Db	1955	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACAAATGAACCTTTCT	2014
Qy	1981	GATTTTGTATTTCATTTATTCCAGCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAG	2040
Db	2015	GATTTTGTATTTCATTTATTCCAGCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAG	2074
Qy	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Db	2075	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2134
Qy	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160
Db	2135	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2194
Qy	2161	GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGATCCAAGCAGGCC	2220
Db	2195	GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGATCCAAGCAGGCC	2254
Qy	2221	TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT	2280
Db	2255	TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT	2314
Qy	2281	TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2340
Db	2315	TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2374

RESULT 5
 AX456520
 LOCUS AX456520 2516 bp DNA linear PAT 06-JUL-2002
 DEFINITION Sequence 42 from Patent WO0227016.
 ACCESSION AX456520
 VERSION AX456520.1 GI:21715410
 KEYWORDS .
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.
 REFERENCE 1
 AUTHORS Patel,S.B. and Dean,M.
 TITLE Gene involved in dietary sterol absorption and excretion and uses therefor
 JOURNAL Patent: WO 0227016-A 42 04-APR-2002;
 THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
 Shailendra B. (US) ; Dean, Michael (US)
 FEATURES Location/Qualifiers
 source 1. .2516
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Primer"

ORIGIN

Query Match 99.9%; Score 2338.4; DB 6; Length 2516;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	GTCAAGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	60
Db	35	GTCAAGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	94
Qy	61	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Db	95	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	154
Qy	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Db	155	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	214
Qy	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	215	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	274
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Db	275	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	334
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	335	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	394
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	395	AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	454

Qy	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	455	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	514
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Db	515	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	574
Qy	541	CGAGACGCTGCACTACACGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	575	CGAGACGCTGCACTACACGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	634
Qy	601	GAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	635	GAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	694
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	695	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	754
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	755	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	814
Qy	781	CATGACTGCTAATCAGATTGTCGTCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	815	CATGACTGCTAATCAGATTGTCGTCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	874
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	875	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	934
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	935	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	994
Qy	961	TGACTGCGGTTACCCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Db	995	TGACTGCGGTTACCCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1054
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1055	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1114
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1115	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1174
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1175	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1234
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1235	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTTACAAGAACTTGGTGAGAAATAAGCTGGC	1294

Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Db	1295	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1354
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCTAGGTCTCCTTTACCA	1380
Db	1355	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCTAGGTCTCCTTTACCA	1414
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1415	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1474
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAAGTGGCAGATGATGCT	1500
Db	1475	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAAGTGGCAGATGATGCT	1534
Qy	1501	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1535	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1594
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCGATTTGGATATTTTCTGCTGC	1620
Db	1595	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCGATTTGGATATTTTCTGCTGC	1654
Qy	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1680
Db	1655	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1714
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG	1740
Db	1715	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG	1774
Qy	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCATTCTTTTAAATCATCAGTTATTT	1800
Db	1775	ATCTGGATTCTCAGAAACATACAAGAAATGCCATTCTTTTAAATCATCAGTTATTT	1834
Qy	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1835	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1894
Qy	1861	CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG	1920
Db	1895	CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG	1954
Qy	1921	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCACAATGAACTTTCT	1980
Db	1955	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCACAATGAACTTTCT	2014
Qy	1981	GATTTTGTATTCAATTTATTCCAGCTCTTGTGATCCTAGGAATAGTTGTTTCAAATAAG	2040
Db	2015	GATTTTGTATTCAATTTATTCCAGCTCTTGTGATCCTAGGAATAGTTGTTTCAAATAAG	2074
Qy	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAAATGGAAGTGAAGCTGCCG	2100
Db	2075	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAAATGGAAGTGAAGCTGCCG	2134
Qy	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160

Db	2135		ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2194
Qy	2161		GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC	2220
Db	2195		GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC	2254
Qy	2221		TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT	2280
Db	2255		TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT	2314
Qy	2281		TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2340
Db	2315		TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2374

RESULT 6

AX456519

LOCUS AX456519 1920 bp DNA linear PAT 06-JUL-2002

DEFINITION Sequence 41 from Patent WO0227016.

ACCESSION AX456519

VERSION AX456519.1 GI:21715409

KEYWORDS .

SOURCE synthetic construct

ORGANISM synthetic construct
artificial sequences.

REFERENCE 1

AUTHORS Patel,S.B. and Dean,M.

TITLE Gene involved in dietary sterol absorption and excretion and uses therefor

JOURNAL Patent: WO 0227016-A 41 04-APR-2002;
THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
Shailendra B. (US) ; Dean, Michael (US)

FEATURES Location/Qualifiers

source 1. .1920
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Primer"

ORIGIN

Query Match 82.1%; Score 1920; DB 6; Length 1920;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	143	ATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCC	202
Db	1	ATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCC	60
Qy	203	CCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGG	262
Db	61	CCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGG	120
Qy	263	CCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTC	322
Db	121	CCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTC	180

Qy	323	TCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAA	382
Db	181	TCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAA	240
Qy	383	ACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAG	442
Db	241	ACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAG	300
Qy	443	GTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTC	502
Db	301	GTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTC	360
Qy	503	CTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCG	562
Db	361	CTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCG	420
Qy	563	CTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATG	622
Db	421	CTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATG	480
Qy	623	GCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGC	682
Db	481	GCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGC	540
Qy	683	ATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAG	742
Db	541	ATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAG	600
Qy	743	GTCATGCTGTTTGATGAGCCAACCACAGGCCCTGGACTGCATGACTGCTAATCAGATTGTC	802
Db	601	GTCATGCTGTTTGATGAGCCAACCACAGGCCCTGGACTGCATGACTGCTAATCAGATTGTC	660
Qy	803	GTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCC	862
Db	661	GTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCC	720
Qy	863	CGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATT	922
Db	721	CGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATT	780
Qy	923	TTCTGTGGCACGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCT	982
Db	781	TTCTGTGGCACGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCT	840
Qy	983	GAACATTCAAACCTTTTGACTTCTATATGGACCTGACGTGAGTGATACCCAAAGCAAG	1042
Db	841	GAACATTCAAACCTTTTGACTTCTATATGGACCTGACGTGAGTGATACCCAAAGCAAG	900
Qy	1043	GAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCA	1102
Db	901	GAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCA	960
Qy	1103	GCAATTTGTCATAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCA	1162
Db	961	GCAATTTGTCATAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCA	1020
Qy	1163	ATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTG	1222

Db	1021	 ATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTCTCTAAACTGGGTGTTCTCCTG	1080
Qy	1223	AGGAGAGTGACAAGAAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAG	1282
Db	1081	 AGGAGAGTGACAAGAAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAG	1140
Qy	1283	AATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTA	1342
Db	1141	 AATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTA	1200
Qy	1343	AAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTAC	1402
Db	1201	 AAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTAC	1260
Qy	1403	ACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAG	1462
Db	1261	 ACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAG	1320
Qy	1463	AGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTC	1522
Db	1321	 AGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTC	1380
Qy	1523	CCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTA	1582
Db	1381	 CCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTA	1440
Qy	1583	CATCCTGAGGTTGCCCCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCTTAATT	1642
Db	1441	 CATCCTGAGGTTGCCCCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCTTAATT	1500
Qy	1643	GGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGT	1702
Db	1501	 GGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGT	1560
Qy	1703	GTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCCCTCAGAAACATA	1762
Db	1561	 GTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCCCTCAGAAACATA	1620
Qy	1763	CAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAAATATTGCAGT	1822
Db	1621	 CAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAAATATTGCAGT	1680
Qy	1823	GAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTT	1882
Db	1681	 GAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTT	1740
Qy	1883	TCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTC AATTCAATTGAGAAAACC	1942
Db	1741	 TCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTC AATTCAATTGAGAAAACC	1800
Qy	1943	TGCCCAGGTGCAACATCTAGATT CACAATGAAC TTTCTGATTTTGTATTCA TTTATTCCA	2002
Db	1801	 TGCCCAGGTGCAACATCTAGATT CACAATGAAC TTTCTGATTTTGTATTCA TTTATTCCA	1860
Qy	2003	GCTCTTGT CATCCTAGGAATAGTTG TTTTCAA AATAAGGGATCATCTCATTAGCAGGTAG	2062

Db 1861 GCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTAG 1920

RESULT 7

AY195873

LOCUS AY195873 2351 bp mRNA linear ROD 01-JUN-2003

DEFINITION Mus musculus strain PERA/Ei ATP-binding cassette sub-family G member 5 (Abcg5) mRNA, complete cds.

ACCESSION AY195873

VERSION AY195873.1 GI:31322257

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2351)

AUTHORS Wittenburg,H., Lyons,M.A., Li,R., Churchill,G.A., Carey,M.C. and Paigen,B.

TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred Mice

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2351)

AUTHORS Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.

TITLE Direct Submission

JOURNAL Submitted (11-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

FEATURES Location/Qualifiers

source 1. .2351

/organism="Mus musculus"

/mol_type="mRNA"

/strain="PERA/Ei"

/db_xref="taxon:10090"

/chromosome="17"

/map="55 cM"

/sex="male"

/tissue_type="liver"

gene 1. .2351

/gene="Abcg5"

CDS 139. .2097

/gene="Abcg5"

/note="ATP-dependent canalicular cholesterol transporter; white subfamily"

/codon_start=1

/product="ATP-binding cassette sub-family G member 5"

/protein_id="AAO45094.1"

/db_xref="GI:31322258"

/translation="MGELPFLSPEGARGPHINRGSLSLSLEQGSVTGTEARHSLGVLHV
SYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISG
RLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS
SADFYNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLD
EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG
TPEEMLGFFNNGCPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESD
IYHKILENIERARYLKTLPVPFKTKDPPGMFGKLGVLRLRRVTRNLNRNKQAVIMRLV
QNLIMGLFLIFYLLRVQNNLTGKAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVS
DQESQDGLYHKWQMLLAYVLHVLPPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALL
APHLIGEFLTLVLLGIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF

TFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSREFTAN
FLILYGFIPALVILGIVIFKVRDYLSR"

ORIGIN

Query Match 60.3%; Score 1410.8; DB 10; Length 2351;
Best Local Similarity 80.4%; Pred. No. 1.4e-296;
Matches 1665; Conservative 0; Mismatches 402; Indels 3; Gaps 1;

```
Qy      25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTGC 84
      || ||| ||||| || | ||| | ||| | ||| ||||| ||||| ||
Db      57 CTCCCATTGGCTCCTCAGTTAAAGTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 116

Qy      85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCGGAGGGTCCAT 144
      ||||| |||| | ||||| ||||| || | | ||| || | || |||
Db     117 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 176

Qy     145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
      || | || | ||||| || | ||||| ||||| || | ||| ||| ||
Db     177 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTGCGTCACGGGCAC 236

Qy     205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAG 261
      ||| || ||||| | || ||| ||| ||||| ||||| ||||| ||| ||
Db     237 AGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTACAGCGTCAGCAACCGTGTCTCG 296

Qy     262 GCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGT 321
      ||| ||||| ||||| || ||| ||| ||||| ||||| ||||| |||||
Db     297 GCCTTGGTGGAACATCAAATCATGCCAGCAGAAAGTGGGACAGGCAAATCCTCAAAGATGT 356

Qy     322 CTCCTTGTTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAA 381
      ||||| |||| | |||| || ||||| ||||| ||||| ||||| |||||
Db     357 CTCCTTGTTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAA 416

Qy     382 AACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGA 441
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     417 GACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGA 476

Qy     442 GGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGT 501
      ||||| ||||| ||| | | ||||| ||| ||||| ||||| ||||| |||||
Db     477 GGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGT 536

Qy     502 CCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGGAGACGCTGCACTACACCGC 561
      ||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||
Db     537 CCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGC 596

Qy     562 GCTGCTGGCCATCCGCCGCGGAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCAT 621
      | ||||| || ||||| || | | ||| | | ||||| ||||| |||||
Db     597 GATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCAT 656

Qy     622 GGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGG 681
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db     657 GACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGG 716

Qy     682 CATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAA 741
      ||||| || ||||| ||||| || ||||| ||||| || ||||| || ||
Db     717 AATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAA 776

Qy     742 GGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGT 801
```

Db	777		836
		GGTCATGATGCTAGATGAGCCAACACAGGACTGGACTGCATGACTGCAAATCAAATTGT	
Qy	802	CGTCCTCCTGGTGGAACCTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCC	861
Db	837	CCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGTCACCATCCACCAGCC	896
Qy	862	CCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGAT	921
Db	897	TCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGT	956
Qy	922	TTTCTGTGGCAGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTC	981
Db	957	GTTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCC	1016
Qy	982	TGAACATTCAAACCCCTTTTGACTTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAA	1041
Db	1017	TGAACATTCCAATCCCTTTGATTCTACATGGACTTGACATCAGTGGACACCCAAAGCAG	1076
Qy	1042	GGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATC	1101
Db	1077	AGAGCGGGAAATAGAAACGTACAAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATC	1136
Qy	1102	AGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACC	1161
Db	1137	TGACATCTATCACAAAATTCTGGAGAACATTGAAAGAGCACGATACCTGAAAACCTTACC	1196
Qy	1162	AATGGTTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTCTCTAAACTGGGTGTTCTCCT	1221
Db	1197	CACGGTTTCCTTTCAAACAAAAGATCCTCCTGGGATGTTGCGCAAGCTTGGTGTCTGCT	1256
Qy	1222	GAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCA	1281
Db	1257	GAGGCGAGTAACAAGAACTTAATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCA	1316
Qy	1282	GAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCT	1341
Db	1317	GAATCTGATCATGGGCCTCTTCCTCATTTTCTACCTTCTCCGCGTCCAGAACAACACGCT	1376
Qy	1342	AAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTA	1401
Db	1377	AAAGGGCGCTGTGCAGGACCGCGTGGGCTGCTCTATCAGCTTGTGGGTGCCACCCATA	1436
Qy	1402	CACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGA	1461
Db	1437	CACCGGCATGCTCAATGCTGTGAATCTGTTTCCCATGCTGAGAGCCGTGAGCGACCAGGA	1496
Qy	1462	GAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCT	1521
Db	1497	GAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCT	1556
Qy	1522	CCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTT	1581
Db	1557	CCCCCTTCAGCGTCATCGCCACGGTCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTT	1616
Qy	1582	ACATCCTGAGGTTGCCCGATTGGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAAT	1641

Db	1617	GTATCCTGAAGTTGCCAGATTTGGATATTTCTCTGCTGCTCTTTTGGCCCCCTCACTTAAT	1676
Qy	1642	TGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAATCCAAATATAGTCAACAG	1701
Db	1677	TGGAGAATTTCTAACACTTGTGCTGCTTGGTATAGTCCAAACCCCTAATATTGTCAACAG	1736
Qy	1702	TGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGGATCTGGATTCCCTCAGAAACAT	1761
Db	1737	TATAGTGGCTCTGCTCAGCATCTCTGGGCTGCTTATTGGATCTGGATTATCAGAAACAT	1796
Qy	1762	ACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAAATATTGCAG	1821
Db	1797	ACAAGAAATGCCCATTCCTTTTAAAATCCTGGGTTATTTTACATTCCAAAAATACTGTTG	1856
Qy	1822	TGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGT	1881
Db	1857	TGAGATTCTCGTGGTCAATGAGTTTTACGGCCTGAACTTCACTTGTGGTGGATCCAACAC	1916
Qy	1882	TTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAAC	1941
Db	1917	CTCTATGCTAAATCACCCGATGTGCGCCATCACCCAAGGGGTCCAGTTCATCGAGAAAAC	1976
Qy	1942	CTGCCCAGGTGCAACATCTAGATTACAAATGAACTTTCTGATTTTGTATTCATTTATTCC	2001
Db	1977	CTGCCCAGGTGCTACATCCAGATTACAGGCAAACCTCCTCATCTTATATGGGTTTATCCC	2036
Qy	2002	AGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTA	2061
Db	2037	AGCTCTGGTCATCCTAGGAATAGTGATTTTTAAAGTCAGGGACTACCTGATTAGCAGATA	2096
Qy	2062	GTGAAAGCCATGGCTGGGAAAATGGAAGTG	2091
Db	2097	GTTAAGATGACAGGCAGGAAAGGGTTAATG	2126

AX456524

DEFINITION Sequence 46 from Patent WO0227016.

ACCESSION AX456524

VERSTON AX456524.1 GI:21715413

KEYWORDS

SOURCE synthetic construct

ORGANISM synthetic construct
 artificial sequences.

REFERENCE 1

AUTHORS Patel, S.B. and Dean, M.

TITLE Gene involved in dietary sterol absorption and excretion and uses therefor

JOURNAL Patent: WO 0227016-A 46 04-APR-2002;
THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
Shailendra B. (US) ; Dean, Michael (US)

FEATURES	Location/Qualifiers
----------	---------------------

```
source      1. .2354
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
```

/note="Primer"

ORIGIN

Query Match 60.2%; Score 1409.2; DB 6; Length 2354;
Best Local Similarity 80.4%; Pred. No. 3.2e-296;
Matches 1664; Conservative 0; Mismatches 403; Indels 3; Gaps 1;

```
Qy      25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
      || ||| ||||| || | ||| | ||| | ||| ||||| ||||| |
Db      57 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 116

Qy      85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT 144
      ||||| ||||| ||| ||||| ||| ||| ||| ||| ||| |||
Db     117 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 176

Qy     145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
      || | || | ||||| || | ||||| ||||| || | ||| |||
Db     177 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCAC 236

Qy     205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAG 261
      ||| || ||||| || ||| ||| ||| ||||| ||||| ||| |||
Db     237 AGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTTACAGCGTCAGCAACCGTGTCGG 296

Qy     262 GCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGT 321
      ||| ||||| ||||| ||| ||| ||| ||| ||||| ||||| |||||
Db     297 GCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGT 356

Qy     322 CTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAA 381
      ||||| ||| ||||| || ||||| ||||| ||||| ||||| |||||
Db     357 CTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAA 416

Qy     382 AACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGA 441
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     417 GACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGA 476

Qy     442 GGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGT 501
      ||||| ||||| ||| | ||||| ||| ||||| ||||| ||||| |||||
Db     477 GGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGT 536

Qy     502 CCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGC 561
      ||||| ||||| | ||||| ||||| ||||| ||| ||||| ||
Db     537 CCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGC 596

Qy     562 GCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCAT 621
      | ||||| || ||||| || | ||| || | ||||| ||||| |||||
Db     597 GATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCAT 656

Qy     622 GGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGG 681
      | ||||| ||||| ||||| ||| ||| | ||||| ||| | |||||
Db     657 GACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGG 716

Qy     682 CATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAA 741
      ||||| || ||||| ||||| || ||||| ||||| || ||||| ||
Db     717 AATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGCAGCCCACTCCTTCAGGACCCCAA 776

Qy     742 GGTGCTGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGT 801
      ||||| || | ||||| ||||| ||||| ||||| ||||| |||||
```

Db 777 GGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGT 836

Qy 802 CGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCC 861
| | ||| ||| || ||||| ||||| ||||| ||||| ||||| |||||

Db 837 CCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGTCACCATCCACCAGCC 896

Qy 862 CCGTTCTGAGCTTTTTGAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGAT 921
|| ||||| || || | ||| ||||| ||||| ||||| ||||| |||||

Db 897 TCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGT 956

Qy 922 TTTCTGTGGCAGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCC 981
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 957 GTTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCC 1016

Qy 982 TGAACATTCAAACCCTTTTGACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAA 1041
||||| || || ||||| || || ||||| ||||| ||||| ||||| |||||

Db 1017 TGAACATTCCAATCCCTTTGATTTTACATGGACTTGACATCAGTGGACACCCAAAGCAG 1076

Qy 1042 GGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATC 1101
|| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

Db 1077 AGAGCGGAAATAGAAACGTACAAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATC 1136

Qy 1102 AGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACC 1161
| || | ||| |||| | || ||||| ||||| ||||| ||||| |||||

Db 1137 TGACATCTATCACAAAATTCTGGAGAACATTGAAAGAGCACGATACCTGAAAACCTTACC 1196

Qy 1162 AATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTCTCTAAACTGGGTGTTCTCCT 1221
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1197 CACGGTTCCTTTCAAACCAAAGATCCTCCTGGGATGTTCCGCAAGCTTGGTGTCTGCT 1256

Qy 1222 GAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCA 1281
|||| |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1257 GAGGCGAGTAACAAGAACTTAATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCA 1316

Qy 1282 GAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCT 1341
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1317 GAATCTGATCATGGGCCTCTTCCTCATTTTCTACCTTCTCCGCGTCCAGAACAACACGCT 1376

Qy 1342 AAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTA 1401
||||| ||| | ||||| ||||| || || || || || ||||| ||||| |||||

Db 1377 AAAGGGCGCTGTGCAGGACCGCGTGGGCTGCTCTATCAGCTTGTGGGTGCCACCCATA 1436

Qy 1402 CACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTCTGCGAGCTGTCAGCGACCAGGA 1461
||| ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1437 CACCGGCATGCTCAATGCTGTGAATCTGTTTCCCATGCTGAGAGCCGTCAGCGACCAGGA 1496

Qy 1462 GAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCT 1521
||||| ||||| || || ||||| ||||| ||||| ||||| ||||| |||||

Db 1497 GAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCT 1556

Qy 1522 CCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTT 1581
||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

Db 1557 CCCCTTCAGCGTCATCGCCACGGTCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTT 1616

Qy 1582 ACATCCTGAGGTTGCCCCGATTTGGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAAT 1641
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1617 GTATCCTGAAGTTGCCAGATTTGGATATTTCTGCTGCTCTTTTGGCCCCCACTTAAT 1676

JOURNAL Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

REFERENCE 3 (bases 1 to 2354)

AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

REMARK Sequence update by submitter

COMMENT On May 16, 2001 this sequence version replaced gi:12382299.

FEATURES

Location/Qualifiers

source 1. .2354

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/tissue_type="liver"

gene 1. .2354

/gene="Abcg5"

CDS 139. .2097

/gene="Abcg5"

/note="ABCG5"

/codon_start=1

/product="sterolin"

/protein_id="AAG53097.1"

/db_xref="GI:12382300"

/translation="MGELPFLSPEGARGPHINRGSLSLSLEQGSVTGTGEARHSLGVLHV
SYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISG
RLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTVRETLYRTAMLALCRS
SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLD
EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG
TPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESD
IYHKILENIERARYLKTLPVFPKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLV
QNLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVS
DQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALL
APHLIGEFLLTVLLGIVQNPINVSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
TFQKYCCIEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN
FLILYGFIPALVILGIVIFKVRDYILIS"

ORIGIN

Query Match 60.2%; Score 1409.2; DB 10; Length 2354;
Best Local Similarity 80.4%; Pred. No. 3.2e-296;
Matches 1664; Conservative 0; Mismatches 403; Indels 3; Gaps 1;

Qy 25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
|| ||| ||||| || | |||| | |||| | || | ||||| ||||| |
Db 57 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 116

Qy 85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCAT 144
||||| |||| | ||||| |||| | | | ||| || | || |||
Db 117 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 176

Qy 145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
|| | || | ||||| || | ||||| |||| | | | ||| || |
Db 177 AGGGCCTCACATCAACAGAGGTCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCAC 236

Qy 205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAG 261
 ||| || | |||| | || ||| ||| ||||| ||||| ||| || |
 Db 237 AGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTCTGG 296

Qy 262 GCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGT 321
 ||| ||||| ||||| ||| ||| ||| ||||| ||||| ||||| |||||
 Db 297 GCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGT 356

Qy 322 CTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAA 381
 ||||| ||| | |||| | |||| | |||| | |||| | |||| | ||||
 Db 357 CTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAA 416

Qy 382 AACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGA 441
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 417 GACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGA 476

Qy 442 GGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGT 501
 ||||| ||||| ||| | | ||||| ||| ||||| ||||| ||||| |||||
 Db 477 GGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGT 536

Qy 502 CCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGC 561
 ||||| ||||| | ||||| ||||| ||||| ||||| ||| ||||| ||
 Db 537 CCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGC 596

Qy 562 GCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCAT 621
 | ||||| || |||| | | | ||| | | ||||| |||| | ||||
 Db 597 GATGCTGGCCCTCTGCCGCGAGCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCAT 656

Qy 622 GGCAGAGCTGAGTCTGAGCCATGTGGCAGACCAGTATTGGCAACTACAGCTTGGGGGG 681
 | ||||| ||||| ||||| |||| | ||| | ||||| ||| | || |||||
 Db 657 GACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGG 716

Qy 682 CATTTCCACGGGTGAGCGGCGCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAA 741
 ||||| || ||||| |||| | ||||| ||||| || || ||||| || ||
 Db 717 AATTTCCAGTGGCGAGCGGCGCGGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAA 776

Qy 742 GGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGT 801
 ||||| || | ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 777 GGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGT 836

Qy 802 CGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACACAGCC 861
 | | ||| || | ||||| ||||| ||||| ||||| || ||||| |||||
 Db 837 CCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGTCACCATCCACAGCC 896

Qy 862 CCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGAT 921
 || ||||| || || | ||| ||||| ||||| ||||| | ||||| || |
 Db 897 TCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGT 956

Qy 922 TTTCTGTGGCACGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCCTGTGCC 981
 ||||| |||| ||| ||||| ||||| |||| | ||||| ||||| |||||
 Db 957 GTTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTGCC 1016

Qy 982 TGAACATTCAAACCTTTTGACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAA 1041
 ||||| || || |||| | || |||| | || ||||| ||||| |||||
 Db 1017 TGAACATTCCAATCCCTTTGATTTTACATGGACTTGACATCAGTGGACACCCAAAGCAG 1076

Qy 1042 GGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATC 1101

Db	1077	AGAGCGGGAAATAGAAAACGTACAAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATC	1136
Qy	1102	AGCAATTTGTCATAAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACC	1161
Db	1137	TGACATCTATCACAAAATTCTGGAGAACATTGAAAGAGCACGATACCTGAAAACCTTACC	1196
Qy	1162	AATGGTTCCTTTCAAAACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCT	1221
Db	1197	CACGGTTCCTTTCAAAACCAAAGATCCTCCTGGGATGTTTCGGCAAGCTTGGTGTCTCTGCT	1256
Qy	1222	GAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCA	1281
Db	1257	GAGGCGAGTAACAAGAACTTAATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCA	1316
Qy	1282	GAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCT	1341
Db	1317	GAATCTGATCATGGGCCTCTTCTCATTTTCTACCTTCTCCGCGTCCAGAACAACACGCT	1376
Qy	1342	AAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTA	1401
Db	1377	AAAGGGCGCTGTGCAGGACCGCGTGGGGCTGCTCTATCAGCTTGTGGGTGCCACCCCAT	1436
Qy	1402	CACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGA	1461
Db	1437	CACCGGCATGCTCAATGCTGTGAATCTGTTTCCCATGCTGAGAGCCGTCAGCGACCAGGA	1496
Qy	1462	GAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACGTCACGTCCT	1521
Db	1497	GAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCT	1556
Qy	1522	CCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTT	1581
Db	1557	CCCCTTCAGCGTCATCGCCACGGTCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTT	1616
Qy	1582	ACATCCTGAGGTTGCCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCTTAAT	1641
Db	1617	GTATCCTGAAGTTGCCAGATTTGGATATTTCTCTGCTGCTCTTTTGGCCCCCTCACTTAAT	1676
Qy	1642	TGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAG	1701
Db	1677	TGGAGAATTTCTAACACTTGCTGCTTGGTATAGTCCAAAACCCTAATATTGTCAACAG	1736
Qy	1702	TGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCTCAGAAACAT	1761
Db	1737	TATAGTGGCTCTGCTCAGCATCTCTGGGCTGCTTATTGGATCTGGATTTATCAGAAACAT	1796
Qy	1762	ACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAATATTGCAG	1821
Db	1797	ACAAGAAATGCCCATTCCTTTTAAAATCCTGGGTTATTTTACATTCCAAAATACTGTTG	1856
Qy	1822	TGAGATTCTTGTAAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGT	1881
Db	1857	TGAGATTCTCGTGGTCAATGAGTTTACGGCCTGAACTTCACTTGTGGTGGATCCAACAC	1916
Qy	1882	TTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAAC	1941

Db	1917	CTCTATGCTAAATCACCCGATGTGCGCCATCACCCAAGGGGTCCAGTTCATCGAGAAAAC	1976
Qy	1942	CTGCCCAGGTGCAACATCTAGATTACAATGAACTTTCTGATTTTGTATTCATTTATTCC	2001
Db	1977	CTGCCCAGGTGCTACATCCAGATTACGGCAAACCTCCTCATCTTATATGGGTTTATCCC	2036
Qy	2002	AGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTA	2061
Db	2037	AGCTCTGGTCATCCTAGGAATAGTGATTTTAAAGTCAGGGACTACCTGATTAGCAGATA	2096
Qy	2062	GTGAAAGCCATGGCTGGGAAAATGGAAGTG	2091
Db	2097	GTTAAGATGACAGGCAGGAAAGGGTTAATG	2126

```

/protein_id="AA045093.1"
/db_xref="GI:31322256"
/translation="MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHV
SYSVSNRVGPWWNIKSCQKQWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISG
RLRCTGTLEGDFVNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS
SADFYNNKKVEAVMTELSLSHVADQVIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLD
EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG
TPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLSAFKESD
IYHKILENIERARYLKTLPVFPKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLV
QNLIMGLFLIFYLLRVQNNLTGKAVQDRVGLLYQFVGATPYTGMLNAVNLFPMRAVS
DQESQDGLYHKWQMLLAYVLHALPFSIIATVIFSSVCYWTGLYPEVARFGYFSAALL
APHLIGEFLTLVLLGIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
TFQKYCCCEILVVNEFYGLNFTCGESNTTMLNHPMCAITQGVFIEKTCPGATSRFTAN
FLILYGFIPALVILGIVIFKVRDYILSR"

```

ORIGIN

```

Query Match          60.1%;  Score 1406;  DB 10;  Length 2351;
Best Local Similarity 80.3%;  Pred. No. 1.6e-295;
Matches 1662;  Conservative    0;  Mismatches 405;  Indels    3;  Gaps    1;

```

```

Qy      25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
      || ||| ||||| || | |||| | |||| | || | ||||| ||||| |||
Db      57 CTCCCATTTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 116

Qy      85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCAT 144
      ||||| |||| | ||||| |||| | || | ||| || | || |||
Db     117 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTCTGAGTCCAGAGGGAGCCAG 176

Qy     145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
      || | || | ||||| || | ||||| ||||| | || | ||| || | |||
Db     177 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTGCGGTACAGGGCAC 236

Qy     205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAG 261
      |||| || ||||| | || |||| |||| | ||||| ||||| |||| || |
Db     237 GGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTACAGCGTCAGCAACCGTGTCGG 296

Qy     262 GCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGT 321
      ||| ||||| ||||| ||| |||| |||| |||| | |||| | ||||| |||||
Db     297 GCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGT 356

Qy     322 CTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAA 381
      ||||| |||| | |||| || |||| | ||||| |||| | ||||| |||||
Db     357 CTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAA 416

Qy     382 AACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCTGGGGGA 441
      ||||| ||||| ||||| ||||| ||||| || || | ||||| | |||||
Db     417 GACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGTGCACTGGGACCCTGGAAGGGGA 476

Qy     442 GGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCTACGT 501
      |||| ||||| ||| | | ||||| |||| | ||||| ||||| ||||| |||||
Db     477 CGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCTACGT 536

Qy     502 CCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGC 561
      ||||| ||||| | ||||| ||||| ||||| |||| | |||| |||
Db     537 CCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGC 596

Qy     562 GCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCAT 621

```

Db	597	GATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCAT	656
Qy	622	GGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGG	681
Db	657	GACAGAGCTGAGCCTGAGCCACGTGGCAGACCAAGTGATTGGCAGCTATAATTTTGGGGG	716
Qy	682	CATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAA	741
Db	717	AATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAA	776
Qy	742	GGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGT	801
Db	777	GGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGT	836
Qy	802	CGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTCCACCAGCC	861
Db	837	CCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGTCACCATCCACCAGCC	896
Qy	862	CCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGAT	921
Db	897	TCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGT	956
Qy	922	TTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAATGACTGCGGTTACCCCTGTCC	981
Db	957	GTTCTGTGGCACCCAGAGGAGATGCTTGGATTCTTCAATAACTGTGGTTACCCCTGTCC	1016
Qy	982	TGAACATTCAAACCTTTTGACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAA	1041
Db	1017	TGAACATTCCAATCCCTTTGATTTCTACATGGACTTGACATCAGTGGACACCCAAAGCAG	1076
Qy	1042	GGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATC	1101
Db	1077	AGAGCGGGAAATAGAAACATACAAGCGAGTACAGATGCTGGAATCTGCCTTCAAGGAATC	1136
Qy	1102	AGCAATTTGTCATAAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACC	1161
Db	1137	TGACATCTATCACAAAATTCTGGAGAACATTGAAAGAGCACGATACCTGAAAACCTTACC	1196
Qy	1162	AATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTCTCTAAACTGGGTGTTCTCCT	1221
Db	1197	CACGGTTCCTTTTAAACAAAAGATCCTCCTGGGATGTTCCGCAAGCTTGGTGTCTCCTGCT	1256
Qy	1222	GAGGAGAGTGACAAGAAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCA	1281
Db	1257	GAGGCGAGTAACAAGAAACTTAATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCA	1316
Qy	1282	GAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCT	1341
Db	1317	GAATCTGATCATGGGCCTCTTCCTCATTTTCTACCTTCTCCGAGTCCAGAACAACACGCT	1376
Qy	1342	AAAGGGTGCTATCCAGGACCGCTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTA	1401
Db	1377	AAAGGGTGCTGTGCAGGACCGCTGGGGCTGCTCTATCAGTTTGTGGGTGCCACCCATA	1436
Qy	1402	CACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGA	1461

Db 1437 CACCGGCATGCTCAATGCTGTGAATCTGTTCCCATGCTGAGAGCCGTCAGCGACCAGGA 1496

Qy 1462 GAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCT 1521
 ||||| |||| | | ||||| |||| | | |||| |

Db 1497 GAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCTGCTCGCCTACGTGCTACACGCCCT 1556

Qy 1522 CCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTT 1581
 ||||| | | |||| | ||||| ||||| |||||

Db 1557 CCCCTTCAGCATCATCGCCACGGTGATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTT 1616

Qy 1582 ACATCCTGAGGTTGCCCCGATTTGGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAAT 1641
 ||||| |||| | ||||| ||||| ||||| |||||

Db 1617 GTATCCTGAAGTTGCCAGATTTGGATATTTCTCTGCTGCTCTTTTGGCCCCCTCACTTAAT 1676

Qy 1642 TGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAG 1701
 || | ||||| ||||| ||||| ||||| ||||| |||||

Db 1677 TGGAGAATTTCTAACACTTGTGCTACTTGGTATAGTCCAAAACCCTAATATTGTCAACAG 1736

Qy 1702 TGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCTCAGAAACAT 1761
 | ||||| |||| | | |||| | ||||| |||||

Db 1737 TATAGTGGCTCTGCTCAGCATCTCTGGACTGCTTATTGGATCTGGATTATCAGAAACAT 1796

Qy 1762 ACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAATATTGCAG 1821
 ||||| ||||| |||| | | ||||| ||||| || |

Db 1797 ACAAGAAATGCCTATTCCTTTTAAAATCCTGGGTATTTTACATTCCAAAATACTGTTG 1856

Qy 1822 TGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGT 1881
 ||||| || | ||||| |||| | |||| | || |

Db 1857 TGAGATTCTCGTGGTCAATGAGTTTACGGCCTGAACCTCACTTGTGGTGAATCCAACAC 1916

Qy 1882 TTCTGTGACAATAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAAC 1941
 || || || | | |||| | |||| | | |||| |

Db 1917 CACTATGCTAAATCACCCGATGTGCGCCATCACCCAAGGGGTCGAGTTCATCGAGAAAAC 1976

Qy 1942 CTGCCCAGGTGCAACATCTAGATTACAAATGAACTTCTGATTTTGTATTCAATTTATTCC 2001
 ||||| |||| | |||| | |||| || || || || || ||

Db 1977 CTGCCCAGGTGCTACATCCAGATTACGGCAAACCTCCTCATCTTATATGGGTTTATCCC 2036

Qy 2002 AGCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTA 2061
 ||||| ||||| ||||| |||| | |||| | || ||||| ||

Db 2037 AGCTCTGGTCATCCTAGGAATAGTGATTTTAAAGTCAGGGACTACCTGATTAGCAGATA 2096

Qy 2062 GTGAAAGCCATGGCTGGGAAAATGGAAGTG 2091
 || || | | |||| | | ||

Db 2097 GTTAAGATGACAGGCAGGAAAGGGTTAATG 2126

RESULT 11

AX320881

LOCUS AX320881 2258 bp DNA linear PAT 14-DEC-2001

DEFINITION Sequence 2 from Patent WO0179272.

ACCESSION AX320881

VERSION AX320881.1 GI:17902431

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Tian,H., Schultz,J. and Shan,B.

TITLE Sitosterolemia susceptibility gene (ssg): compositions and methods of use

JOURNAL Patent: WO 0179272-A 2 25-OCT-2001; Tularik Inc. (US)

FEATURES Location/Qualifiers

source 1. .2258
 /organism="Mus musculus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:10090"
 /note="mouse sitosterolemia susceptibility gene (SSG)"

CDS 47. .2005
 /note="unnamed protein product; mouse sitosterolemia susceptibility gene (SSG) protein"
 /codon_start=1
 /protein_id="CAD19408.1"
 /db_xref="GI:17902432"
 /db_xref="REMTREMBL:CAD19408"
 /translation="MGELPFLSPEGARGPHINRGSLSSEQGSVTGTEARHSLGVLHV
 SYSVSNRVGPWWNIKSCQKQKWDQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISG
 RLRRRTGTLEGEVFVNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLLRYTAMLALCRS
 SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVISIAAQLQDPKVMMLD
 EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG
 TPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESD
 IYHKILENIERARYLKTLPMPVFKTKDPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLV
 QNLIMGLFLIFYLLRVQNNTLKGAQVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVS
 DQESQDGLYHKWQMLLAYVLHVLPPFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALL
 APHLIGEFLTLVLLGIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
 TFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN
 FLILYGFIPALVILGIVIFKVRDYILISR"

ORIGIN

Query Match 59.6%; Score 1395.6; DB 6; Length 2258;
 Best Local Similarity 80.7%; Pred. No. 3e-293;
 Matches 1642; Conservative 0; Mismatches 389; Indels 3; Gaps 1;

Qy 61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
 ||| | ||||| ||||| | ||||| ||| | ||||| || |

Db 1 GGGACAGGCCACTAGAAAATTCATTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCC 60

Qy 121 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
 | | ||| || | || || | || | || | ||||| || | ||||| ||

Db 61 CTTTCTGAGTCCAGAGGGAGCCAGAGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCT 120

Qy 181 GGAGGGGGCTCCTGCCACCGCCCCGAGCCT---CACAGCCTGGGCATCCTCCATGCCTC 237
 |||| | | | | || | | | || | ||||| | | ||||| ||

Db 121 GGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTC 180

Qy 238 CTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTG 297
 ||||| ||||| |||| | |||| | |||| | |||| | |||| | ||||

Db 181 CTACAGCGTCAGCAACCGTGTGGGCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTG 240

Qy 298 GACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCAT 357
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 241 GGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCAT 300

Qy	358	CCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGG	417
Db	301	CTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCG	360
Qy	418	GCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCA	477
Db	361	GCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCA	420
Qy	478	GTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGT	537
Db	421	GTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGT	480
Qy	538	GCGCGAGACGCTGCACTACACGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTT	597
Db	481	GCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTA	540
Qy	598	CCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACT	657
Db	541	CAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAT	600
Qy	658	GATTGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGC	717
Db	601	GATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGC	660
Qy	718	AGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGA	777
Db	661	AGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGA	720
Qy	778	CTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAAT	837
Db	721	CTGCATGACTGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAAT	780
Qy	838	TGTGGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGC	897
Db	781	TGTGATTGTCACCATCCACAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGC	840
Qy	898	CATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTT	957
Db	841	CATCCTGACTTACGGAGAGTTGGTGTTCGTGGCACCCAGAGGAGATGCTTGGCTTCTT	900
Qy	958	CAATGACTGCGGTTACCCCTTGTCCTGAACATTCAAACCCTTTTGACTTCTATATGGACCT	1017
Db	901	CAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTGATTTTTACATGGACTT	960
Qy	1018	GACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGAT	1077
Db	961	GACATCAGTGGACACCCAAAGCAGAGAGCGGGAATAGAAACGTACAAGCGAGTACAGAT	1020
Qy	1078	GATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAG	1137
Db	1021	GCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATCTGGAGAACATTGAAAG	1080
Qy	1138	AATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGT	1197
Db	1081	AGCACGATACCTGAAAACCTTACCCATGGTTCCTTTCAAACCAAAGATCCTCCTGGGAT	1140

Qy 1198 TTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCT 1257
 ||| || || |||| || |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1141 GTTCGGCAAGCTTGGTGTCTGCTGAGGCGAGTAACAAGAACTTAATGAGGAATAAGCA 1200

Qy 1258 GGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGT 1317
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1201 GGCAGTGATTATGCGTCTCGTTCAGAATCTGATCATGGGCCTCTTCCTCATTTCCTACCT 1260

Qy 1318 TCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTA 1377
 ||| || |||| | ||| |||| |||| ||| || |||| |||| || || || || ||
 Db 1261 TCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGCGTGGGGCTGCTCTA 1320

Qy 1378 CCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGT 1437
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1321 TCAGCTTGTGGGTGCCACCCATACACGGCATGCTCAATGCTGTGAATCTGTTTCCCAT 1380

Qy 1438 GCTGCGAGCTGTGAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGAT 1497
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1381 GCTGAGAGCCGTGAGCGACCAGGAGAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCT 1440

Qy 1498 GCTGGCCTATGCACTGCACGTCCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAG 1557
 ||| |||| | || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1441 GCTCGCCTACGTGCTACACGTCCTCCCTTCAGCGTCATCGCCACGGTCATTTTCAGCAG 1500

Qy 1558 TGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTTCTGC 1617
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1501 TGTGTGTTATTGGACTCTGGGCTTGTATCCTGAAGTTGCCAGATTTGGATATTTCTCTGC 1560

Qy 1618 TGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGT 1677
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1561 TGCTCTTTTGGCCCCCTCACTTAATTGGAGAATTTCTAACACTTGTGCTGCTTGGTATAGT 1620

Qy 1678 CCAAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGT 1737
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1621 CCAAAACCCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATCTCTGGGCTGCTTAT 1680

Qy 1738 TGGATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTA 1797
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1681 TGGATCTGGATTTATCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCCTGGGTTA 1740

Qy 1798 TTTTACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAA 1857
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1741 TTTTACATTCCAAAAATACTGTTGTGAGATTCTCGTGGTCAATGAGTTTACGGCCTGAA 1800

Qy 1858 TTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCA 1917
 |||| |||| || || || || || || || || || || || || || || || || || || ||
 Db 1801 CTTCACCTTGTGGTGGATCCAACACCTCTATGCTAAATCACCCGATGTGCGCCATCACCCA 1860

Qy 1918 AGGAATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTT 1977
 ||| | || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1861 AGGGGTCCAGTTCATCGAGAAAACCTGCCCAGGTGCTACATCCAGATTCACGGCAAACCTT 1920

Qy 1978 TCTGATTTTGTATTTCATTTATTCCAGCTCTTGTATCCTAGGAATAGTTGTTTTCAAAAT 2037
 || || || || || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1921 CCTCATCTTATATGGGTTTATCCCAGCTCTGGTCATCCTAGGAATAGTGATTTTAAAGT 1980

Qy 2038 AAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTG 2091

Db 1981 CAGGGACTACCTGATTAGCAGATAGTTAAGATGACAGGCAGGAAAGGGTTAATG 2034

RESULT 12

AF312714

LOCUS AF312714 2470 bp mRNA linear ROD 26-AUG-2002

DEFINITION Rattus norvegicus sterolin (Abcg5) mRNA, complete cds.

ACCESSION AF312714

VERSION AF312714.3 GI:22477143

KEYWORDS .

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 2470)

AUTHORS Lee,M.H., Lu,K., Hazard,S., Yu,H., Shulenin,S., Hidaka,H.,
Kojima,H., Allikmets,R., Sakuma,N., Pegoraro,R., Srivastava,A.K.,
Salen,G., Dean,M. and Patel,S.B.

TITLE Identification of a gene, ABCG5, important in the regulation of
dietary cholesterol absorption

JOURNAL Nat. Genet. 27 (1), 79-83 (2001)

MEDLINE 20578753

PUBMED 11138003

REFERENCE 2 (bases 1 to 2470)

AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403, USA

REFERENCE 3 (bases 1 to 2470)

AUTHORS Lu,K., Lee,M.-H. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403, USA

REMARK Sequence update by submitter

REFERENCE 4 (bases 1 to 2470)

AUTHORS Lu,K., Lee,M. and Patel,S.B.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-2002) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St, STB 541, Charleston, SC 29403, USA

REMARK Sequence update by submitter

COMMENT On Aug 26, 2002 this sequence version replaced gi:14091945.

FEATURES Location/Qualifiers

source 1. .2470
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="liver"
gene 1. .2470
/gene="Abcg5"
CDS 65. .2023
/gene="Abcg5"

```

/note="ABCG5"
/codon_start=1
/product="sterolin"
/protein_id="AAG53098.3"
/db_xref="GI:22477144"
/translation="MSELPFLSPEGARGPHNNRGSQSSLEEGSVTGSEARHSLGVLNV
SFSVSNRVGPWWNIKSCQQKWDRKILKDVSLYIESGQTMCILGSSSGSGKTLLDAISG
RLRRTGTLEGEVFNVCCELRRDQFQDCVSYLLQSDVFLSSLTVRETLRYTAMLALRSS
SADFYDKKVEAVLTELSLSHVADQMIGNYNFNGGISSGERRRVSIAAQLQDPKVMMLD
EPTTGLDCMTANHIVLLLVELARRNRIVIVTIHQPRSELFHHFDKIAILTYGELVFCG
TPEEMLGFFNNGCYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLESAFRQSD
ICHKILENIERTRLKTLPMVPFKTKNPPGMFCKLGVLLRRVTRNLMRNKQVVMRLV
QNLIMGLFLIFYLLRVQNNMLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVS
DQESQDGLYQKWQMLLAYVLHALPFSIVATVIFSSVCYWTGLGYPEVARFGYFSAALL
APHLIGEFLTLVLLGMVQNPNIIVNSIVALLSISGLLIGSGFIRNIEEMPIPLKILGYF
TFQKYCCEILVVNEFYGLNFTCGGSNTSVPNPNMCSMTQGIQFIEKTCPGATSRTTN
FLILYSFIPTLVILGMVVFVKVRDYLISR"

```

ORIGIN

```

Query Match          59.1%;   Score 1383.8;   DB 10;   Length 2470;
Best Local Similarity 80.0%;   Pred. No. 1.1e-290;
Matches 1641;   Conservative    0;   Mismatches 407;   Indels    3;   Gaps    1;

```

```

Qy      44 TGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTG 103
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2 TAAAGTTGCTCTGAAGCCAGACAGGACACCAGAGGATTCACATTCCTCCCGCTG 61

Qy     104 GCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGA 163
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      62 GCCATGAGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACAACAACAGA 121

Qy     164 GGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTG 220
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     122 GGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTCAGAGGCTCGGCACAGCTTA 181

Qy     221 GGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACA 280
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     182 GGTGTCCTGAATGTGTCTTCAGCGTCAGCAACCGTGTGGGGCCCTGGTGAACATCAAA 241

Qy     281 TCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGC 340
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     242 TCATGCCAGCAGAAGTGGGACAGGAAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGT 301

Qy     341 GGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCC 400
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     302 GGCCAGACCATGTGCATCTTAGGTAGCTCAGGCTCAGGGAAAACCACGCTGCTGGACGCC 361

Qy     401 ATGTCCGGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGG 460
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     362 ATCTCTGGGAGGCTGCGGCGCACAGGGACCTTGAAGGGGAAGTGTGTTGTGAACGGCTGC 421

Qy     461 GCGCTGCGCCGGGAGCAGTTCAGGACTGCTTCTCCTACGTCTGTCAGAGCGACACCCTG 520
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     422 GAGCTGCGCAGGGACAGTTCGAAGACTGCGTCTCCTACCTCCTGTCAGAGCGATGTCTTT 481

Qy     521 CTGAGCAGCCTCACCGTGCGGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGC 580
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	482	CTGAGCAGCCTCACGGTGC	GGGAGACGCTGAGATACACGGCGATGCTGGCTCTCCGCAGC	541
Qy	581	GGCAATCCCGGCTCCTTCC	AAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGC	640
Db	542	AGCTCCGCGGACTTCTACG	ACAAGAAGGTAGAGGCAGTCCTGACAGAGCTGAGTCTGAGC	601
Qy	641	CATGTGGCAGACCGACTG	ATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGG	700
Db	602	CACGTGGCAGACCAAATG	ATCGGCAACTATAATTTTGGGGGGATTTCAGTGGCGAGCGG	661
Qy	701	CGCCGGGTCTCCATCGC	AGCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAG	760
Db	662	CGCCGAGTGTCCATCGC	AGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTTGACGAG	721
Qy	761	CCAACCACAGGCCTGG	ACTGTCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGA	820
Db	722	CCAACCACAGGACTGG	ACTGTCATGACTGCAAAATCATATCGTCCTCCTCTGGTCGAG	781
Qy	821	GCTCGCAGGAACCGA	ATTGTGGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAG	880
Db	782	GCTCGCAGGAACCGC	ATTGTAATTGTCACCATCCACAGCCTCGCTCTGAGCTCTTCCAC	841
Qy	881	CTCTTTGACAAAATTG	CCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCG	940
Db	842	CAC TTCGACAAAATTG	CCATTCTGACTTACGGAGAGTTGGTGTCTGTGGCAGCCAGAG	901
Qy	941	GAAATGCTTGATTTC	TCTCAATGACTGCGGTTACCCTTGTCTGAACATTCAAACCCCTTTT	1000
Db	902	GAGATGCTCGGCTTCT	TCAATAACTGTGGTTACCCTTGTCTGAACATTCCAATCCCTTT	961
Qy	1001	GACTTCTATATGGAC	CTGACGTCA GTGGATACCCAAAGCAAGGAACGGGAAATAGAAACC	1060
Db	962	GATTTCTACATGGAC	TTGACATCGGTGGACACCCAAAGCAGAGAGCGAGAGATAGAGACG	1021
Qy	1061	TCCAAGAGAGTCCAG	ATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAAACT	1120
Db	1022	TACAAGCGAGTCCAG	ATGCTGGAATCTGCCTTCAGGCAATCGGACATCTGTACAAAATC	1081
Qy	1121	TTGAAGAATATTGAA	AGAAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAA	1180
Db	1082	CTGGAGAACATTGAA	AGAACAAGACACCTGAAAACCCCTACCCATGGTTCCTTTCAA	1141
Qy	1181	AAAGATTCTCCTGG	AGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAAAC	1240
Db	1142	AAAAATCCTCCCGG	AATGTTCTGCAAGCTCGGCGTTCTCCTGAGGAGAGTAACGAGAAAC	1201
Qy	1241	TTGGTGAGAAATAAG	CTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTG	1300
Db	1202	CTAATGAGGAATAAG	CAGGTGGTGATTATGCGTCTTGTTTCAGAATCTGATCATGGGTTTG	1261
Qy	1301	TTCCTCCTTTTCTTC	GTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGAC	1360
Db	1262	TTCCTCATTTTCTAC	CTTCTCCGAGTCCAGAACAACATGCTGAAGGGCGCTGTTTCAGGAC	1321
Qy	1361	CGCGTAGGTCTCCTT	TACCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCT	1420
Db	1322	CGCGTAGGGCTGTTG	TACCAGCTTGTGGGTGCCACCCCGTACACCGGCATGCTCAACGCT	1381

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.
 TITLE Abcg5 and abcg8: compositions and methods of use
 JOURNAL Patent: WO 02081691-A 1 17-OCT-2002;
 Tularik Inc. (US) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
 (US)

FEATURES Location/Qualifiers
 source 1. .1959
 /organism="Mus musculus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:10090"
 CDS 1. .1959
 /note="unnamed protein product; ABCG5 (mABCG5)"
 /codon_start=1
 /protein_id="CAD86570.1"
 /db_xref="GI:29371739"
 /db_xref="REMTREMBL:CAD86570"
 /translation="MGELPFLSPEGARGPHINRGSLSLSLEQGSVTGTEARHSLGVLHV
 SYSVSNRVGPWWNIKSCQKQKWDRIKDVSLYIESGQIMCILGSSSGSKTTLDDAISG
 RLRRGTGLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS
 SADFYNNKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLDPKVMMLD
 EPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFDKIAILTYGELVFCG
 TPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESD
 IYHKILENIERARYLKTLPMPVFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLV
 QNLIMGLFLFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVS
 DQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALL
 APHLIGEFLTLVLLGIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYF
 TFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN
 FLILYGFIPALVILGIVIFKVRDYLISR"

ORIGIN

Query Match 58.4%; Score 1365.4; DB 6; Length 1959;
 Best Local Similarity 81.4%; Pred. No. 1.1e-286;
 Matches 1595; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

Qy 107 ATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGC 166
 ||||| || | | | || | | | || | | | |||||
 Db 1 ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 60
 Qy 167 TCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTGGGGC 223
 || | ||||| ||||| | | | | || | | | ||||| ||
 Db 61 TCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120
 Qy 224 ATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCT 283
 |||| |||| | ||||| ||||| || || |||| | ||||| |||
 Db 121 GTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTTCGGGCCTTGGTGGAAACATCAATCA 180
 Qy 284 TGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGG 343
 |||| |||| |||| | |||| | ||||| ||||| ||||| ||||| ||
 Db 181 TGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGC 240
 Qy 344 CAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATG 403
 |||| | ||||| |||| | ||||| |||| | ||||| ||||| ||||| |||||

Db	241	CAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATC	300
Qy	404	TCCGGGAGGCTGGGGCGCGCGGGGACCTTCTGGGGGAGGTGTATGTGAACGGCCGGGCG	463
Db	301	TCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAG	360
Qy	464	CTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTG	523
Db	361	CTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTG	420
Qy	524	AGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGC	583
Db	421	AGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGCGAGC	480
Qy	584	AATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCAT	643
Db	481	TCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCAC	540
Qy	644	GTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGC	703
Db	541	GTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGCGGAGCGGCGC	600
Qy	704	CGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCA	763
Db	601	CGAGTTTCCATCGCAGCCCACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCA	660
Qy	764	ACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACCTGGCT	823
Db	661	ACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCT	720
Qy	824	CGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTCAGCTC	883
Db	721	CGCAGGGACCGAATTGTGATTGTCACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACAC	780
Qy	884	TTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAA	943
Db	781	TTTGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCAGAGGAG	840
Qy	944	ATGCTTGATTTCTTCAATGACTGCGGTTACCCTTGTCTGAACATTCAAACCTTTTGAC	1003
Db	841	ATGCTTGGCTTCTTCAATAACTGTGGTTACCCTGTCTGAACATTCCAATCCCTTTGAT	900
Qy	1004	TTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCC	1063
Db	901	TTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTAC	960
Qy	1064	AAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAAACTTTG	1123
Db	961	AAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATCTGACATCTATCAGAAATTTCTG	1020
Qy	1124	AAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAA	1183
Db	1021	GAGAACATTGAAAGAGCACGATACCTGAAAACCTTACCCATGGTTCCTTTCAAACAAAA	1080
Qy	1184	GATTCTCCTGGAGTTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTG	1243
Db	1081	GATCCTCCTGGGATGTTTCGGCAAGCTTGGTGTCTCCTGCTGAGGCGAGTAACAAGAACTTA	1140

[illegible]

AX456526

DEFINITION Sequence 48 from Patent WO0227016.

VERSION AX456526.1 GI:21715414

KEYWORDS

SOURCE synthetic construct

ORGANISM synthetic construct
artificial sequences.

REFERENCE 1

AUTHORS Patel, S.B. and Dean, M.

TITLE Gene involved in dietary sterol absorption and excretion and uses therefor

JOURNAL Patent: WO 0227016-A 48 04-APR-2002;

THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel, Shailendra B. (US) ; Dean, Michael (US)

FEATURES	Location/Qualifiers
----------	---------------------

```
source      1. .2035
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
            /note="primer"
```

Query Match 58.2%; Score 1363; DB 6; Length 2035;

Best Local Similarity 80.6%; Pred. No. 3.7e-286;

Matches 1607; Conservative 0; Mismatches 385; Indels 3; Gaps 1;

Qy 100 GTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAA 159
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1 GCTGGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACAACAA 60

Qy 160 CAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCGGAGCCT---CACAG 216

Db 61 CAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTCAGAGGCTCGGCACAG 120

QY 217 CCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACAT 276

Db 121 CTTAGGTGTCCTGAATGTGTCCTTCAGCGTCAGCAACCGTGTCGGGCCCTGGTGGAAACAT 180

QY 277 CACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGA 336
|||

Db 181 CAAATCATGCCAGCAGAAGTGGGACAGGAAAATCCTCAAAGATGTCTCCTTGTACATCGA 240

Qy 337 GAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAAACCACGCTGCTGGA 396
|||

Db 241 GAGTGGCCAGACCATGTGCATCTTAGGTAGCTCAGGCTCAGGGAAAACCACGCTGCTGGA 300

Qy 397 CGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGG 456

Db 301 CGCCATCTCTGGGAGGCTGCGGCGCACAGGGACCTTGGAAGGGGAAGTGTTTGTGAACGG 360

Qy 457 CCGGGCGCTGCGCCGGGAGCAGTTCAGGACTGCTTCTCCTACGTCTGCAGAGCGACAC 516

Db 361 CTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCGTCTCCTACCTCCTGCAGAGCGATGT 420

Qy	517	CCTGCTGAGCAGCCTCACCGTGCCGCAGACGCTGCACTACACCGCGCTGCTGGCCATCCG	576
Db	421	CTTTCTGAGCAGCCTCACGGTGC GG GAGACGCTGAGATACACGGCGATGCTGGCTCTCCG	480
Qy	577	CCGCGGCAATCCC GG CTCTTCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCT	636
Db	481	CAGCAGCTCCGCGGACTTCTACGACAAGAAGGTAGAGGCAGTCCTGACAGAGCTGAGTCT	540
Qy	637	GAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGA	696
Db	541	GAGCCACGTGGCAGACC AATGATCGGCAACTATAATTTGGGGGGATTTCCAGTGGCGA	600
Qy	697	GCGGCGCCGGGTCTCCATCGCAGCCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGA	756
Db	601	GCGGCGCCGAGTGTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTTGA	660
Qy	757	TGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGA	816
Db	661	CGAGCCAACCACAGGACTGGACTGCATGACTGCAAATCATATCGTCCTCCTCTTGGTCGA	720
Qy	817	ACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTT	876
Db	721	GCTGGCTCGCAGGAACCGCATTGTAATTGT CACC ATCCACAGCCTCGCTCTGAGCTCTT	780
Qy	877	TCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCC	936
Db	781	CCACCACTTCGACAAAATTGCCATTCTGACTTACGGAGAGTTGGTGT TCTGTGGCACGCC	840
Qy	937	AGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCCTGTCTGAACATTCAAACCC	996
Db	841	AGAGGAGATGCTCGGCTTCTTCAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCC	900
Qy	997	TTTTGACTTCTATATGGACCTGACGT CAG TG GATACCCAAAGCAAGGAACGGGAAATAGA	1056
Db	901	CTTTGATTTCTACATGGACTTGACATCGGTGGACACCCAAAGCAGAGAGCGAGAGATAGA	960
Qy	1057	AACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAA	1116
Db	961	GACGTACAAGCGAGTCCAGATGCTGGAATCTGCCTTCAGGCAATCGGACATCTGTCACAA	1020
Qy	1117	AACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAA	1176
Db	1021	AATCCTGGAGAACATTGAAAGAACAAGACACCTGAAAACCCATCCCATGGTTCCTTTCAA	1080
Qy	1177	AACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAG	1236
Db	1081	AACGAAAAATCCTCCCGGAATGTTCTGCAAGCTCGGCGTTCTCCTGAGGAGAGTAACGAG	1140
Qy	1237	AAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGG	1296
Db	1141	AAACCTAATGAGGAATAAGCAGGTGGTGATTATGCGTCTTGTT CAG AATCTGATCATGGG	1200
Qy	1297	TTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCA	1356
Db	1201	TCTGTTCTCATTTTTCTACCTTCTCCGAGTCCAGAACACATGCTGAAGGGCGCTGTTCA	1260

Qy	1357	GGACCGCGTAGGTCTCTTTTACCAGTTTGTGGGGGCCACCCCGTACACAGGCATGCTGAA	1416
Db	1261	GGACCGCGTAGGGCTGTTGTACCAGCTTGTGGGTGCCACCCCGTACACCGGCATGCTCAA	1320
Qy	1417	CGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCT	1476
Db	1321	CGCTGTGAACCTCTTTCCCATGCTGAGAGCTGTCAGCGACCAGGAGAGTCAGGATGGCCT	1380
Qy	1477	CTACCAGAAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCTCCCTTCAGCGTTGT	1536
Db	1381	GTACCAGAAAGTGGCAGATGCTGCTCGCCTATGTGCTGCATGCTCTCCCTTCAGCATCGT	1440
Qy	1537	TGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGC	1596
Db	1441	TGCCACGGTGATTTTCAGCAGCGTGTGTTACTGGACTCTGGGCTTGTATCCCGAGGTCGC	1500
Qy	1597	CCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAAC	1656
Db	1501	CAGATTTGGATACTTCTCTGCCGCTCTGTTGGCCCCCTACTTAATTGGAGAATTTCTGAC	1560
Qy	1657	TCTTGTGCTACTTGGTATCGTCCAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCT	1716
Db	1561	ACTTGTGCTGCTTGGTATGGTCCAAACCCCAATATTGTCAACAGCATAGTGGCTCTGCT	1620
Qy	1717	GTCCATTGCGGGGTGCTTGTTGGATCTGGATTCCCTCAGAAACATACAAGAAATGCCCAT	1776
Db	1621	GAGTATTTCTGGGTGCTCATTGGATCTGGATTTATCAGAAACATAGAAGAAATGCCCAT	1680
Qy	1777	TCCTTTTAAATCATCAGTTATTTTACATTCCAAAAATATTGCAGTGAGATTCTTGTAGT	1836
Db	1681	TCCTTTTAAAAATCCTGGGTTACTTTACCTTCCAAAAGTACTGTTGTGAGATTCTTGTGGT	1740
Qy	1837	CAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAA	1896
Db	1741	CAATGAGTTCTATGGCCTGAACCTCACTTGTGGTGGCTCCAACACTTCTGTGCCAAATAA	1800
Qy	1897	TCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAACCTGCCAGGTGCAAC	1956
Db	1801	CCCAATGTGTTCCATGACCCAAGGGATCCAATTCATTGAGAAAACCTGCCAGGGGCCAC	1860
Qy	1957	ATCTAGATTACAAATGAACTTTCTGATTTTGTATTCAATTTATTCCAGCTCTTGTATCCT	2016
Db	1861	GTCCAGATTACAGACAAACTTCTGATCTTGTACTCGTTCATCCCGACTCTTGTATCCT	1920
Qy	2017	AGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCT	2076
Db	1921	GGGGATGGTGGTCTTTAAAGTCCGGGACTACCTGATTAGCAGATAGGTAAGATGGCAGGC	1980
Qy	2077	GGGAAAATGGAAGTG	2091
Db	1981	AGGAAAGGGTTAATG	1995

AX456523

LOCUS	AX456523	1915 bp	DNA	linear	PAT 06-JUL-2002
DEFINITION	Sequence 45 from Patent WO0227016.				

ACCESSION AX456523
 VERSION AX456523.1 GI:21715412
 KEYWORDS .
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.
 REFERENCE 1
 AUTHORS Patel,S.B. and Dean,M.
 TITLE Gene involved in dietary sterol absorption and excretion and uses therefor
 JOURNAL Patent: WO 0227016-A 45 04-APR-2002;
 THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
 Shailendra B. (US) ; Dean, Michael (US)

FEATURES Location/Qualifiers
 source 1. .1915
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Primer"

ORIGIN

Query Match 57.1%; Score 1335.8; DB 6; Length 1915;
 Best Local Similarity 81.5%; Pred. No. 3.1e-280;
 Matches 1560; Conservative 0; Mismatches 352; Indels 3; Gaps 1;

Qy	107	ATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCATGGGTCTCCAAGTAAACAGAGGC	166
Db	1	ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG	60
Qy	167	TCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGAGCCT---CACAGCCTGGGC	223
Db	61	TCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT	120
Qy	224	ATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCT	283
Db	121	GTCTTGCATGTGTCCTACAGCGTCAGCAACCGTGTCTGGGCTTGGTGGAAACATCAAATCA	180
Qy	284	TGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGG	343
Db	181	TGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGC	240
Qy	344	CAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATG	403
Db	241	CAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATC	300
Qy	404	TCCGGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCG	463
Db	301	TCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTGTGAATGGCTGCGAG	360
Qy	464	CTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTG	523
Db	361	CTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTG	420
Qy	524	AGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGC	583
Db	421	AGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTTGCCGCAGC	480

Qy	584	AATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCAT	643
Db	481	TCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCAC	540
Qy	644	GTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGC	703
Db	541	GTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCAGCGGCGC	600
Qy	704	CGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCA	763
Db	601	CGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCA	660
Qy	764	ACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCTCCTCCTGGTGGAACTGGCT	823
Db	661	ACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCTTCTCTTGCTGAGCTGGCT	720
Qy	824	CGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTTCAGCTC	883
Db	721	CGCAGGGACCGAATTGTGATTGTACCATTCCACCAGCCTCGCTCTGAGCTCTTCCAACAC	780
Qy	884	TTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAA	943
Db	781	TTTGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCCAGAGGAG	840
Qy	944	ATGCTTGATTCTTCAATGACTGCGGTTACCCCTTGTCTGAACATTCAAACCCTTTTGAC	1003
Db	841	ATGCTTGCTTCTTCAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTGAT	900
Qy	1004	TTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCC	1063
Db	901	TTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTAC	960
Qy	1064	AAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAAACTTTG	1123
Db	961	AAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTG	1020
Qy	1124	AAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAA	1183
Db	1021	GAGAACATTGAAAGAGCACGATACCTGAAAACCTTACCCACGGTTCCTTTCAAACAAAA	1080
Qy	1184	GATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTG	1243
Db	1081	GATCCTCCTGGGATGTTTCGGCAAGCTTGGTGTCTGCTGAGGCGAGTAACAAGAACTTA	1140
Qy	1244	GTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTT	1303
Db	1141	ATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTTCAAGATCTGATCATGGGCTCTTC	1200
Qy	1304	CTCCTTTTCTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGC	1363
Db	1201	CTCATTTTCTACCTTCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGC	1260
Qy	1364	GTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCGTACACAGGCATGCTGAACGCTGTG	1423
Db	1261	GTGGGGCTGCTCTATCAGCTTGTGGGTGCCACCCATACACGGGCATGCTCAATGCTGTG	1320
Qy	1424	AATCTGTTTCCCGTGTCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAG	1483

Db	1321	AATCTGTTTCCCATGCTGAGAGCCGTGAGAGCCAGGAGAGTCAGGATGGCCTGTATCAT	1380
Qy	1484	AAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACC	1543
Db	1381	AAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCTCCCCTTCAGCGTCATCGCCACG	1440
Qy	1544	ATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCCGATT	1603
Db	1441	GTCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTTGTATCCTGAAGTTGCCAGATT	1500
Qy	1604	GGATATTTTCTGCTGCTCTCTTGCCCCCCTTAATTGGTGAATTTCTAACTCTTGTG	1663
Db	1501	GGATATTTCTCTGCTGCTCTTTTGCCCCCTCACTTAATTGGAGAATTTCTAACACTTGTG	1560
Qy	1664	CTACTTGGTATCGTCCAAAATCCAAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATT	1723
Db	1561	CTGCTTGGTATAGTCCAAAACCCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATC	1620
Qy	1724	GCGGGGGTGCTTGTGGATCTGGATTCCCTCAGAAACATACAAGAAATGCCCATTCCTTTT	1783
Db	1621	TCTGGGCTGCTTATTGGATCTGGATTATCAGAAACATACAAGAAATGCCCATTCCTTTA	1680
Qy	1784	AAAATCATCAGTTATTTTACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAG	1843
Db	1681	AAAATCCTGGGTATTTTACATTCCAAAAATACTGTTGTGAGATTCTCGTGGTCAATGAG	1740
Qy	1844	TTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAATAATCCAATG	1903
Db	1741	TTTTACGGCCTGAACCTTCACTTGTGGTGGATCCAACACCTCTATGCTAAATCACCCGATG	1800
Qy	1904	TGTGCCTTCACTCAAGGAATTCATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGA	1963
Db	1801	TGCGCCATCACCCAAGGGGTCCAGTTCATCGAGAAAACCTGCCAGGTGCTACATCCAGA	1860
Qy	1964	TTCACAATGAACTTTCTGATTTTGTATTTCATTTATTCAGCTCTTGTATCCTAG	2018
Db	1861	TTCACGGCAAACCTTCCTCATCTTATATGGGTTTATCCCAGCTCTGGTCATCCTAG	1915

Search completed: February 26, 2004, 06:21:17
Job time : 6017.48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:39:18 ; Search time 592.658 Seconds
(without alignments)
16773.223 Million cell updates/sec

Title: US-09-989-981A-5
Perfect score: 2340
Sequence: 1 gtcaggtggagcaggcaggg.....aatattcataaacctatggg 2340

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		DB	ID	Description
	Score	Match Length			
1	2340	100.0 2340	6	AAD22009	Aad22009 Human sit
2	2340	100.0 2340	7	AAD48882	Aad48882 Human ABC
3	2338.4	99.9 2516	6	ABK51682	Abk51682 Human ABC
4	1920	82.1 1920	6	ABK51681	Abk51681 DNA encod
5	1409.2	60.2 2354	6	ABK51685	Abk51685 Mouse ABC
6	1395.6	59.6 2258	6	AAD22008	Aad22008 Mouse sit
7	1365.4	58.4 1959	7	AAD48880	Aad48880 Mouse ABC

	8	1363	58.2	2035	6	ABK51686	Abk51686	cDNA enco
	9	1335.8	57.1	1915	6	ABK51684	Abk51684	DNA encod
	10	1174.2	50.2	2512	9	ADB62671	Adb62671	Human cDN
	11	770.2	32.9	1069	6	ABK51687	Abk51687	cDNA enco
	12	472	20.2	472	6	AAD22022	Aad22022	Human sit
	13	429.4	18.4	432	3	AAC76065	Aac76065	Human ORF
	14	268	11.5	281	3	AAZ94755	Aaz94755	Human ATP
	15	249.6	10.7	5460	6	ABK51683	Abk51683	Human ABC
	16	249	10.6	249	6	AAD22010	Aad22010	Human sit
	17	214	9.1	214	6	AAD22017	Aad22017	Human sit
	18	206	8.8	206	6	AAD22018	Aad22018	Human sit
	19	203.6	8.7	2669	7	AAD48883	Aad48883	Human ABC
	20	199.2	8.5	2019	7	AAD48881	Aad48881	Mouse ABC
	21	199.2	8.5	2564	6	ABN90022	Abn90022	Mouse clo
	22	186	7.9	186	6	AAD22020	Aad22020	Human sit
	23	140	6.0	140	6	AAD22015	Aad22015	Human sit
	24	139	5.9	139	6	AAD22019	Aad22019	Human sit
	25	137	5.9	137	6	AAD22012	Aad22012	Human sit
	26	135.4	5.8	472	8	ACH43425	Ach43425	Human foe
	27	133	5.7	246	6	ABN24693	Abn24693	Human ORF
	28	130	5.6	130	6	AAD22016	Aad22016	Human sit
	29	129	5.5	129	6	AAD22014	Aad22014	Human sit
	30	126.2	5.4	2525	3	AAZ98625	Aaz98625	Silkworm
	31	125.4	5.4	3586	6	ABI99363	Abi99363	Mouse isc
c	32	122.2	5.2	371	7	ABT22948	Abt22948	Breast ca
	33	122	5.2	122	6	AAD22011	Aad22011	Human sit
	34	115.4	4.9	133	2	AAT21044	Aat21044	Human gen
	35	114.8	4.9	2352	4	ABL05135	Abl05135	Drosophil
	36	114.2	4.9	2133	9	ADE47651	Ade47651	Human NOV
	37	114.2	4.9	2894	7	ACD13444	Acd13444	Human DNA
	38	114.2	4.9	2921	7	ABV75074	Abv75074	Human Dev
	39	114.2	4.9	2930	3	AAZ94747	Aaz94747	Human ATP
	40	114.2	4.9	2930	6	ABL63321	Abl63321	Breast ca
	41	114.2	4.9	3201	6	ABV74352	Abv74352	Human ABC
	42	113	4.8	113	6	AAD22021	Aad22021	Human sit
	43	112.4	4.8	2429	8	AAL62515	Aal62515	Human tra
	44	110.8	4.7	2687	6	AAD46413	Aad46413	Human ABC
	45	110.8	4.7	2687	7	ABV77265	Abv77265	Nucleotid

ALIGNMENTS

RESULT 1

AAD22009

ID AAD22009 standard; DNA; 2340 BP.

XX

AC AAD22009;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG).

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21; ds.

```

XX
OS   Homo sapiens.
XX
FH   Key                Location/Qualifiers
FT   CDS                107..2062
FT                               /*tag= a
FT                               /product= "Human SSG protein"
XX
PN   WO200179272-A2.
XX
PD   25-OCT-2001.
XX
PF   18-APR-2001; 2001WO-US012758.
XX
PR   18-APR-2000; 2000US-0198465P.
PR   15-MAY-2000; 2000US-0204234P.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Tian H,  Schultz J,  Shan B;
XX
DR   WPI; 2002-017598/02.
DR   P-PSDB; AAE13290.
XX
PT   Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
PT   useful for screening a compound that increases the level of expression or
PT   activity of SSG polypeptide for treating sterol-related disorder.
XX
PS   Claim 8; Fig 8; 105pp; English.
XX
CC   The invention relates to an isolated Sitosterolaemia Susceptibility Gene
CC   (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC   binding cassette (ABC) family cholesterol transporter. SSG is useful for
CC   identifying a compound useful in the treatment or prevention of a sterol-
CC   related disorder, including sitosterolaemia, hyperlipidaemia,
CC   hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC   nutritional deficiencies. SSG is also useful for treating cholesterol-
CC   associated diseases or conditions including coronary heart disease and
CC   other cardiovascular diseases, and sitosterolaemia-associated condition
CC   including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC   expression cassette is useful in the production of transgenic non-human
CC   animals. SSG genes and their homologues are useful as tools for a number
CC   of applications including diagnosing sitosterolaemia and other
CC   cardiovascular disorders, for forensics and paternity determinations, and
CC   for treating any of a large number of SSG associated diseases. The
CC   present sequence is human SSG DNA. Human SSG is located on chromosome
CC   2p21
XX
SQ   Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

Query Match                100.0%;  Score 2340;  DB 6;  Length 2340;
Best Local Similarity      100.0%;  Pred. No. 0;
Matches 2340;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy                1  GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db                1  GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60

```


Qy	61	GGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Db	61	GGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Qy	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Db	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Qy	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Db	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Qy	421	CGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	421	CGCGGGGACCTTCTTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Db	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Qy	601	GAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	601	GAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Qy	781	CATGACTGCTAATCAGATTGTTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	781	CATGACTGCTAATCAGATTGTTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900

Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Qy	961	TGACTGCGGTTACCCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Db	961	TGACTGCGGTTACCCCTTGTCTGAACATTCAAACCTTTTGACTTCTATATGGACCTGAC	1020
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTCTTCGTTCT	1320
Db	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTCTTCGTTCT	1320
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Db	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTCCCGTGCT	1440
Db	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTCCCGTGCT	1440
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Qy	1501	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1501	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCGATTTGGATATTTTCTGCTGC	1620
Db	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCGATTTGGATATTTTCTGCTGC	1620
Qy	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1680
Db	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1680
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG	1740
Db	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGG	1740
Qy	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT	1800

Db	1741	ATCTGGATTCCCTCAGAAAACATACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTT	1800
Qy	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Qy	1861	CAC TTGTGGCAGCTCAAATGTTTCTGTGACAAC TAATCCAATGTGTGCCTTCACTCAAGG	1920
Db	1861	CAC TTGTGGCAGCTCAAATGTTTCTGTGACAAC TAATCCAATGTGTGCCTTCACTCAAGG	1920
Qy	1921	AATTC AATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCT	1980
Db	1921	AATTC AATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCT	1980
Qy	1981	GATTTTGTATTCA TTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAG	2040
Db	1981	GATTTTGTATTCA TTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAG	2040
Qy	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Db	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Qy	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160
Db	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160
Qy	2161	GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC	2220
Db	2161	GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC	2220
Qy	2221	TTGAATGCAATGGAAGTG GTTTATAGTCCCTTGCTCTTACAAC TTGCAGGGACATGTGGT	2280
Db	2221	TTGAATGCAATGGAAGTG GTTTATAGTCCCTTGCTCTTACAAC TTGCAGGGACATGTGGT	2280
Qy	2281	TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2340
Db	2281	TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2340

RESULT 2

AAD48882

ID AAD48882 standard; DNA; 2340 BP.

XX

AC AAD48882;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG5 DNA.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
KW ABCG5; gene; ds.

XX

OS Homo sapiens.

```

XX
FH Key Location/Qualifiers
FT CDS 107. .2062
FT /*tag= a
FT /product= "hABCG5 protein"
XX
PN WO200281691-A2.
XX
PD 17-OCT-2002.
XX
PF 20-NOV-2001; 2001WO-US043823.
XX
PR 20-NOV-2000; 2000US-0252235P.
PR 28-NOV-2000; 2000US-0253645P.
XX
PA (TULA-) TULARIK INC.
PA (TEXA ) UNIV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tian H;
XX
DR WPI; 2003-058548/05.
DR P-PSDB; AAE31704.
XX
PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
PT related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies.
XX
PS Claim 11; Page 77; 94pp; English.
XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is human ABCG5 DNA
XX
SQ Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 U; 0 Other;

Query Match 100.0%; Score 2340; DB 7; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
Db 1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60

QY 61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
Db 61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120

QY 121 ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCT 180
Db 121 ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCT 180

```

Qy	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Db	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Qy	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGCGCTGCGCCGGGAGCAGTT	480
Db	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGCGCTGCGCCGGGAGCAGTT	480
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Db	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Qy	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGT	840
Db	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGT	840
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Qy	961	TGACTGCGGTTACCCCTTGTCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1020
Db	961	TGACTGCGGTTACCCCTTGTCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1020
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080

Db	1021	 GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCCTCTTTCTTCGTTCT	1320
Db	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCCTCTTTCTTCGTTCT	1320
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Db	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Qy	1501	GGCCTATGCACTGCACGTCCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1501	GGCCTATGCACTGCACGTCCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCCGATTGGATATTTTCTGCTGC	1620
Db	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCCGATTGGATATTTTCTGCTGC	1620
Qy	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1680
Db	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCA	1680
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG	1740
Db	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG	1740
Qy	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT	1800
Db	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT	1800
Qy	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Qy	1861	CACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGG	1920

Db 1861 CACTTGTGGCAGCTCAAATGTTTCTGTGACAACCTAATCCAATGTGTGCCTTCACTCAAGG 1920
 Qy 1921 AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCT 1980
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1921 AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCT 1980
 Qy 1981 GATTTTGTATTCAATTTATTCCAGCTCTTGTCACTCCTAGGAATAGTTGTTTTCAAATAAG 2040
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1981 GATTTTGTATTCAATTTATTCCAGCTCTTGTCACTCCTAGGAATAGTTGTTTTCAAATAAG 2040
 Qy 2041 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2100
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 2041 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2100
 Qy 2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160
 Qy 2161 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 2161 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
 Qy 2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
 Qy 2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340

RESULT 3

ABK51682

ID ABK51682 standard; cDNA; 2516 BP.

XX

AC ABK51682;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 cDNA sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW chromosome 2p21; ss.

XX

OS Homo sapiens.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

XX

XX

XX

XX

XX

XX

10

Query Match 99.9%; Score 2338.4; DB 6; Length 2516;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	60
Db	35	GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA	94
Qy	61	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Db	95	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	154
Qy	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Db	155	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	214
Qy	181	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	215	GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	274
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300

Db	275	 CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	334
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	335	 CAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGGGCAGATCATGTGCATCCT	394
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	395	 AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	454
Qy	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	455	 CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	514
Qy	481	CCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Db	515	 CCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	574
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	575	 CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	634
Qy	601	GAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	635	 GAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	694
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	695	 TGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	754
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	755	 CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	814
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	815	 CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	874
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	875	 GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	934
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	935	 CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATTTCTTCAA	994
Qy	961	TGACTGCGGTTACCCTTGTCCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1020
Db	995	 TGACTGCGGTTACCCTTGTCCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1054
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1055	 GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1114
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140

Db 1115 AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT, 1174

Qy 1141 GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT 1200
 |||

Db 1175 GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT 1234

Qy 1201 CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC 1260
 |||

Db 1235 CTCTAAACTGGGTGTTCTCCTGAGGAGAGTTACAAGAACTTGGTGAGAAATAAGCTGGC 1294

Qy 1261 AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT 1320
 |||

Db 1295 AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT 1354

Qy 1321 GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA 1380
 |||

Db 1355 GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA 1414

Qy 1381 GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT 1440
 |||

Db 1415 GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT 1474

Qy 1441 GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT 1500
 |||

Db 1475 GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT 1534

Qy 1501 GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT 1560
 |||

Db 1535 GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT 1594

Qy 1561 GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTCTGCTGC 1620
 |||

Db 1595 GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTCTGCTGC 1654

Qy 1621 TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA 1680
 |||

Db 1655 TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA 1714

Qy 1681 AAATCCAAATATAGTCAACAGTGAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG 1740
 |||

Db 1715 AAATCCAAATATAGTCAACAGTGAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG 1774

Qy 1741 ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT 1800
 |||

Db 1775 ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTT 1834

Qy 1801 TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT 1860
 |||

Db 1835 TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT 1894

Qy 1861 CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG 1920
 |||

Db 1895 CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG 1954

Qy 1921 AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTACAATGAACCTTCT 1980
 |||

Db 1955 AATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTACAATGAACCTTCT 2014

Qy 1981 GATTTTGTATTTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAATAAG 2040
 |||
 Db 2015 GATTTTGTATTTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAATAAG 2074
 |||
 Qy 2041 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2100
 |||
 Db 2075 GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG 2134
 |||
 Qy 2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160
 |||
 Db 2135 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2194
 |||
 Qy 2161 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
 |||
 Db 2195 GACATCTCAAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2254
 |||
 Qy 2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACTGCAGGGACATGTGGT 2280
 |||
 Db 2255 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACTGCAGGGACATGTGGT 2314
 |||
 Qy 2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
 |||
 Db 2315 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2374
 |||

RESULT 4

ABK51681

ID ABK51681 standard; DNA; 1920 BP.

XX

AC ABK51681;

XX

DT 30-JUL-2002 (first entry)

XX

DE DNA encoding human ABCG5 protein.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW chromosome 2p21; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1. .1920

FT /*tag= a

FT /product= "Human ABCG5 protein"

FT /transl_except= (pos: 4. .9, aa: GDLSSLTPGGSMGL)

FT /note= "This sequence contains 13 exons"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.

XX
PI Patel SB, Dean M;

XX
DR WPI; 2002-416483/44.
DR P-PSDB; AAU98984.

XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.

XX
PS Claim 38; Page 36-37; 66pp; English.

XX
CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the human ABCG5 gene located on chromosome 2p21.
CC This sequence encodes the human ABCG5 protein of the invention

XX
SQ Sequence 1920 BP; 440 A; 503 C; 486 G; 491 T; 0 U; 0 Other;

Query Match 82.1%; Score 1920; DB 6; Length 1920;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

Db 181 TCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAA 240

Qy 383 ACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAG 442
|||||

Db 241 ACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAG 300

Qy 443 GTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTC 502
|||||

Db 301 GTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTC 360

Qy 503 CTGCAGAGCGACACCTTGCTGAGCAGCCTCACCCTGCGCGAGACGCTGCACTACACCGCG 562
|||||

Db 361 CTGCAGAGCGACACCTTGCTGAGCAGCCTCACCCTGCGCGAGACGCTGCACTACACCGCG 420

Qy 563 CTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATG 622
|||||

Db 421 CTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATG 480

Qy 623 GCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGC 682
|||||

Db 481 GCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGC 540

Qy 683 ATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAG 742
|||||

Db 541 ATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAG 600

Qy 743 GTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTC 802
|||||

Db 601 GTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTC 660

Qy 803 GTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCC 862
|||||

Db 661 GTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCC 720

Qy 863 CGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATT 922
|||||

Db 721 CGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATT 780

Qy 923 TTCTGTGGCAGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCT 982
|||||

Db 781 TTCTGTGGCAGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCTTGTCT 840

Qy 983 GAACATTCAAACCTTTTGACTTCTATATGGACCTGACGTCAGTGGATAACCAAAGCAAG 1042
|||||

Db 841 GAACATTCAAACCTTTTGACTTCTATATGGACCTGACGTCAGTGGATAACCAAAGCAAG 900

Qy 1043 GAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCA 1102
|||||

Db 901 GAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCA 960

Qy 1103 GCAATTTGTCATAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCA 1162
|||||

Db 961 GCAATTTGTCATAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCA 1020

Qy 1163 ATGGTTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTCTCTAACTGGGTGTTCTCCTG 1222
|||||

Db 1021 ATGGTTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTCTCTAACTGGGTGTTCTCCTG 1080

Qy	1223	AGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAG	1282
Db	1081	AGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAG	1140
Qy	1283	AATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTA	1342
Db	1141	AATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTA	1200
Qy	1343	AAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTAC	1402
Db	1201	AAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTAC	1260
Qy	1403	ACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAG	1462
Db	1261	ACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAG	1320
Qy	1463	AGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCTC	1522
Db	1321	AGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCTC	1380
Qy	1523	CCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTA	1582
Db	1381	CCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTA	1440
Qy	1583	CATCCTGAGGTTGCCCGATTTGGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAATT	1642
Db	1441	CATCCTGAGGTTGCCCGATTTGGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAATT	1500
Qy	1643	GGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGT	1702
Db	1501	GGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGT	1560
Qy	1703	GTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCCTCAGAAACATA	1762
Db	1561	GTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCCTCAGAAACATA	1620
Qy	1763	CAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTTTACATTCCAAAAATATTGCAGT	1822
Db	1621	CAAGAAATGCCCATTCCTTTTAAATCATCAGTTATTTTACATTCCAAAAATATTGCAGT	1680
Qy	1823	GAGATTCCTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTT	1882
Db	1681	GAGATTCCTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTT	1740
Qy	1883	TCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAACC	1942
Db	1741	TCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAACC	1800
Qy	1943	TGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCTGATTTTGTATTCAATTATTCCA	2002
Db	1801	TGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCTGATTTTGTATTCAATTATTCCA	1860
Qy	2003	GCTCTTGTTCATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTAG	2062
Db	1861	GCTCTTGTTCATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTAG	1920

RESULT 5

ABK51685

ID ABK51685 standard; cDNA; 2354 BP.

XX

AC ABK51685;

XX

DT 30-JUL-2002 (first entry)

XX

DE Mouse ABCG5 cDNA sequence.

XX

KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW ss.

XX

OS Mus sp.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 45; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol

CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of the
CC invention

XX

SQ Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 U; 0 Other;

Query Match 60.2%; Score 1409.2; DB 6; Length 2354;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 1664; Conservative 0; Mismatches 403; Indels 3; Gaps 1;

```
Qy      25 CTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGAAAATTTGC 84
      || ||| ||||| || | |||| | |||| | ||| ||||| ||||| |
Db      57 CTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGAAAATTCAC 116

Qy      85 CCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCAT 144
      ||||| |||| | ||||| |||| | | ||| || | || |||
Db     117 TTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAG 176

Qy     145 GGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCC 204
      || | || | ||||| || | ||||| |||| | | | | || | | |
Db     177 AGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTGCGTCACGGGCAC 236

Qy     205 GGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAG 261
      ||| || ||||| || || |||| |||| ||||| ||||| ||||| |||
Db     237 AGAGGCTCGGCACAGCTTAGGTGTCTGCATGTGTCTACAGCGTCAGCAACCGTGTCTCG 296

Qy     262 GCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGT 321
      ||| ||||| ||||| || |||| |||| ||||| ||||| ||||| |||||
Db     297 GCCTTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGT 356

Qy     322 CTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAA 381
      ||||| |||| | |||| || ||||| ||||| ||||| ||||| |||||
Db     357 CTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAA 416

Qy     382 AACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGA 441
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     417 GACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACTGGGACCTTGAAGGGGA 476

Qy     442 GGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGT 501
      ||||| ||||| ||| | | ||||| |||| ||||| ||||| ||||| |||||
Db     477 GGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGT 536

Qy     502 CCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGGAGACGCTGCACTACACCGC 561
      ||||| ||||| | ||||| ||||| ||||| ||||| ||| |||| ||
Db     537 CCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGGAGACGTTGCGATACACAGC 596

Qy     562 GCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCAT 621
      | ||||| || ||||| || | | ||| | | ||||| ||||| |||||
Db     597 GATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCAT 656

Qy     622 GGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGG 681
      | ||||| ||||| ||||| |||| | ||||| |||| | || |||||
Db     657 GACAGAGCTGAGCCTGAGCCACGTGGCGGACCAATGATTGGCAGCTATAATTTTGGGGG 716

Qy     682 CATTTCCACGGGTGAGCGGCGCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAA 741
      ||||| || ||||| |||| | ||||| ||||| || || ||||| || |
Db     717 AATTTCCAGTGGCGAGCGGCGCGGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAA 776
```


Qy	742	GGTCATGCTGTTTGATGAGCCAACCACAGGCCCTGGACTGCATGACTGCTAATCAGATTGT	801
Db	777	GGTCATGATGCTAGATGAGCCAACCACAGGACTGGACTGCATGACTGCAAAATCAAATTGT	836
Qy	802	CGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACACAGCC	861
Db	837	CCTTCTCTTGGCTGAGCTGGCTCGCAGGACCGAATTGTGATTGTCACCATCCACAGCC	896
Qy	862	CCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGAT	921
Db	897	TCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGT	956
Qy	922	TTTCTGTGGCACGCCAGCGGAAATGCTTGATTCTTCAATGACTGCGGTTACCCCTTGTC	981
Db	957	GTTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTTGTC	1016
Qy	982	TGAACATTCAAACCCCTTTTGACTTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAA	1041
Db	1017	TGAACATTCCAATCCCTTTGATTTTACATGGACTTGACATCAGTGGACACCCAAAGCAG	1076
Qy	1042	GGAACGGGAAATAGAAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATC	1101
Db	1077	AGAGCGGGAAATAGAAAACGTACAAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATC	1136
Qy	1102	AGCAATTTGTCTATAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACC	1161
Db	1137	TGACATCTATCACAAAATTCTGGAGAACATTGAAAGAGCACGATACCTGAAAACCTTACC	1196
Qy	1162	AATGGTTCCTTTCAAAACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCT	1221
Db	1197	CACGGTTCCTTTCAAAACCAAAGATCCTCCTGGGATGTTCCGCAAGCTTGGTGTCTCCT	1256
Qy	1222	GAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCA	1281
Db	1257	GAGGCGAGTAACAAGAACTTAATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCA	1316
Qy	1282	GAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCT	1341
Db	1317	GAATCTGATCATGGGCCTCTTCCTCATTTTCTACCTTCTCCGCGTCCAGAACAACACGCT	1376
Qy	1342	AAAGGGTGCTATCCAGGACCGCTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTA	1401
Db	1377	AAAGGGCGCTGTGCAGGACCGCTGGGGCTGCTCTATCAGCTTGTGGGTGCCACCCCAT	1436
Qy	1402	CACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGA	1461
Db	1437	CACCGGCATGCTCAATGCTGTGAATCTGTTTCCCATGCTGAGAGCCGTGAGCGACCAGGA	1496
Qy	1462	GAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCT	1521
Db	1497	GAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCT	1556
Qy	1522	CCCTTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTTGCTACTGGACGCTGGGCTT	1581
Db	1557	CCCTTTCAGCGTCATCGCCACGGTCATTTTCAGCAGTGTTGTTATTGGACTCTGGGCTT	1616

Qy	1582	ACATCCTGAGGTTGCCCCGATTGGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAAT	1641
Db	1617	GTATCCTGAAGTTGCCAGATTGGATATTTCTCTGCTGCTCTTTGGCCCCCACTTAAT	1676
Qy	1642	TGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAG	1701
Db	1677	TGGAGAATTTCTAACACTTGTGCTGCTTGGTATAGTCCAAAACCCTAATATTGTCAACAG	1736
Qy	1702	TGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGGTGGATCTGGATTCCCTCAGAAACAT	1761
Db	1737	TATAGTGGCTCTGCTCAGCATCTCTGGGCTGCTTATTGGATCTGGATTTATCAGAAACAT	1796
Qy	1762	ACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAATATTGCAG	1821
Db	1797	ACAAGAAATGCCCATTCCTTTTAAAATCCTGGGTTATTTTACATTCCAAAATACTGTTG	1856
Qy	1822	TGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGT	1881
Db	1857	TGAGATTCTCGTGGTCAATGAGTTTACGGCCTGAACCTTCACTTGTGGTGGATCCAACAC	1916
Qy	1882	TTCTGTGACAATAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAAC	1941
Db	1917	CTCTATGCTAAATCACCGATGTGCGCCATCACCCAAGGGGTCCAGTTCATCGAGAAAAC	1976
Qy	1942	CTGCCCAGGTGCAACATCTAGATTACAAATGAACCTTCTGATTTTGTATTCAATTTATTC	2001
Db	1977	CTGCCCAGGTGCTACATCCAGATTACGGCAAACCTTCCTCATCTTATATGGGTTTATCCC	2036
Qy	2002	AGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGTA	2061
Db	2037	AGCTCTGGTGCATCCTAGGAATAGTGATTTTAAAGTCAGGGACTACCTGATTAGCAGATA	2096
Qy	2062	GTGAAAGCCATGGCTGGGAAAATGGAAGTG	2091
Db	2097	GTTAAGATGACAGGCAGGAAAGGGTTAATG	2126

RESULT 6

AAD22008

ID AAD22008 standard; DNA; 2258 BP.

XX

AC AAD22008;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG).

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT CDS 47. .2005

FT /*tag= a

FT /product= "Mouse SSG protein"
 XX
 PN WO200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001WO-US012758.
 XX
 PR 18-APR-2000; 2000US-0198465P.
 PR 15-MAY-2000; 2000US-0204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 XX
 DR WPI; 2002-017598/02.
 DR P-PSDB; AAE13289.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
 PT useful for screening a compound that increases the level of expression or
 PT activity of SSG polypeptide for treating sterol-related disorder.
 XX
 PS Claim 8; Fig 7; 105pp; English.
 XX
 CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
 CC identifying a compound useful in the treatment or prevention of a sterol-
 CC related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations, and
 CC for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG DNA. Mouse SSG is located on chromosome 17
 XX
 SQ Sequence 2258 BP; 549 A; 579 C; 567 G; 563 T; 0 U; 0 Other;

Query Match 59.6%; Score 1395.6; DB 6; Length 2258;
 Best Local Similarity 80.7%; Pred. No. 0;
 Matches 1642; Conservative 0; Mismatches 389; Indels 3; Gaps 1;

Qy 61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
 ||| | ||||| ||||| | ||||| ||| | ||||| ||| |
 Db 1 GGGACAGGCCACTAGAAAATTCATTGCAATTTGCTTCCTGCTAGCCATGGGTGAGCTGCC 60
 Qy 121 ATCTTTGACCCCCGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCAGAGCTCCCT 180
 | | ||| || | || ||| || | || | ||||| || | ||||| |||
 Db 61 CTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCT 120
 Qy 181 GGAGGGGGCTCCTGCCACCGCCCCGAGCCT---CACAGCCTGGGCATCCTCCATGCCTC 237
 |||| | | | ||| | | ||| || ||||| | || |||| ||| ||

Db	121	GGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCTGTCATGTGTCTC	180
Qy	238	CTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTG	297
Db	181	CTACAGCGTCAGCAACCGTGTCTGGGCCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTG	240
Qy	298	GACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCAT	357
Db	241	GGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCAT	300
Qy	358	CCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGG	417
Db	301	CTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCG	360
Qy	418	GCGCGCGGGGACCTTCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCA	477
Db	361	GCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCA	420
Qy	478	GTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGT	537
Db	421	GTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGT	480
Qy	538	GCGCGAGACGCTGCACTACACGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTT	597
Db	481	GCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTA	540
Qy	598	CCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACT	657
Db	541	CAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAAT	600
Qy	658	GATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGC	717
Db	601	GATTGGCAGCTATAATTTTGGGGGAATTTCAGTGGCGAGCGGCGCCGAGTTTCCATCGC	660
Qy	718	AGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGA	777
Db	661	AGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGA	720
Qy	778	CTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAAT	837
Db	721	CTGCATGACTGCAAATCAAATTGTCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAAT	780
Qy	838	TGTGGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGC	897
Db	781	TGTGATTGTACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGC	840
Qy	898	CATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTT	957
Db	841	CATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCACAGAGAGATGCTTGCTTCTT	900
Qy	958	CAATGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTTTGACTTCTATATGGACCT	1017
Db	901	CAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTGATTTTACATGGACTT	960
Qy	1018	GACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGAT	1077
Db	961	GACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTACAAGCGAGTACAGAT	1020

Qy 1078 GATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAG 1137
 | | |||| |||| |||| |||| | | | ||| |||| | | |||| |||||
 Db 1021 GCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTGGAGAACATTGAAAG 1080

Qy 1138 AATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGT 1197
 | | |||| |||| |||| |||| |||| |||| |||| |||| |||| |
 Db 1081 AGCACGATACCTGAAAACCTTACCCATGGTTCCTTTCAAACCAAAGATCCTCCTGGGAT 1140

Qy 1198 TTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCT 1257
 ||| || || |||| || |||| || ||| |||| |||| |||| ||||
 Db 1141 GTTCGGCAAGCTTGGTGTCTGCTGAGGCGAGTAACAAGAACTTAATGAGGAATAAGCA 1200

Qy 1258 GGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGT 1317
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1201 GGCAGTGATTATGCGTCTCGTTCAGAATCTGATCATGGGCTCTTCCTCATTTTCTACCT 1260

Qy 1318 TCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTA 1377
 ||| || |||| | ||| |||| |||| ||| | |||| |||| || || || ||
 Db 1261 TCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGCGTGGGGCTGCTCTA 1320

Qy 1378 CCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGT 1437
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1321 TCAGCTTGTGGGTGCCACCCCATACACCGGCATGCTCAATGCTGTGAATCTGTTTCCCAT 1380

Qy 1438 GCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGAT 1497
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1381 GCTGAGAGCCGTACAGCGACCAGGAGAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCT 1440

Qy 1498 GCTGGCCTATGCACTGCACGTCTCCTCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAG 1557
 ||| |||| | || |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1441 GCTCGCTACGTGCTACACGTCTCCTCCCCTTCAGCGTCATCGCCACGGTCATTTTCAGCAG 1500

Qy 1558 TGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTGGATATTTTCTGC 1617
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1501 TGTGTGTTATTGGACTCTGGGCTTGTATCCTGAAGTTGCCAGATTGGATATTTCTCTGC 1560

Qy 1618 TGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGT 1677
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1561 TGCTCTTTTGGCCCCCACTTAATTGGAGAATTTCTAACACTTGTGCTGCTTGGTATAGT 1620

Qy 1678 CCAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTCTTGT 1737
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1621 CCAAACCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATCTCTGGGCTGCTTAT 1680

Qy 1738 TGGATCTGGATTCCCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTA 1797
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1681 TGGATCTGGATTATCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCCTGGGTTA 1740

Qy 1798 TTTTACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAA 1857
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1741 TTTTACATTCCAAAAATACTGTTGTGAGATTCTCGTGGTCAATGAGTTTACGGCCTGAA 1800

Qy 1858 TTTCACCTTGTGGCAGCTCAAATGTTTCTGTGACAATAATCCAATGTGTGCCCTTCACTCA 1917
 |||| |||| || || || || || || || || || || || || || || ||
 Db 1801 CTTCACCTTGTGGTGGATCCAACACCTCTATGCTAAATCACCCGATGTGCGCCATCACCCA 1860

QY 1918 AGGAATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTCACAATGAACTT 1977
 ||| | || |||| |||||||||||||||||| |||| |||||| ||||
 Db 1861 AGGGGTCCAGTTCATCGAGAAAACCTGCCAGGTGCTACATCCAGATTCACGGCAAACCTT 1920
 QY 1978 TCTGATTTTGTATTCAATTTATCCAGCTCTGTGCATCCTAGGAATAGTTGTTTTCAAAT 2037
 || || || || || |||| |||||| |||||||||| |||| |||| |
 Db 1921 CCTCATCTTATATGGGTTTATCCCAGCTCTGGTCATCCTAGGAATAGTGATTTTTAAAGT 1980
 QY 2038 AAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTG 2091
 |||| | || |||||| |||| || | | |||| | | ||
 Db 1981 CAGGGACTACCTGATTAGCAGATAGTTAAGATGACAGGCAGGAAAGGGTTAATG 2034

RESULT 7

AAD48880

ID AAD48880 standard; DNA; 1959 BP.

XX

AC AAD48880;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG5 DNA.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT CDS 1. .1591

FT /*tag= a

FT /product= "mABCG5 protein"

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

DR P-PSDB; AAE31702.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.

XX

PS Claim 11; Page 73; 94pp; English.

XX

CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is mouse ABCG5 DNA

XX

SQ Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 U; 0 Other;

Query Match 58.4%; Score 1365.4; DB 7; Length 1959;
Best Local Similarity 81.4%; Pred. No. 0;
Matches 1595; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

```
Qy      107 ATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCATGGGTCTCCAAGTAAACAGAGGC 166
          ||||| || | | || | | | || | | | || | | | |||
Db       1 ATGGGTGAGCTGCCCTTTCTGAGTCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 60

Qy      167 TCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGAGCCT---CACAGCCTGGGC 223
          || | ||||| ||||| | | || | | | || | | | |||| | ||
Db       61 TCTCTGAGCTCCCTGGAGCAAGGTTCCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120

Qy      224 ATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCT 283
          |||| |||| | ||||| ||||| || | |||| | |||| | |||
Db      121 GTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTCGGGCCTTGGTGGAAACATCAAATCA 180

Qy      284 TGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGG 343
          |||| |||| |||| | |||| | ||||| ||||| ||||| || | |||| ||
Db      181 TGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGC 240

Qy      344 CAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATG 403
          |||| | ||||| || | ||||| ||||| ||||| ||||| ||||| |||
Db      241 CAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATC 300

Qy      404 TCCGGGAGGCTGGGGCGCGCGGGGACCTTCTGGGGGAGGTGTATGTGAACGGCCGGGCG 463
          ||||| |||| | |||| | | ||||| ||||| |||| | || | || |
Db      301 TCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAG 360

Qy      464 CTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTG 523
          ||||| |||| | ||||| | ||||| ||||| ||||| ||||| || | |||
Db      361 CTGCGCAGGGACAGTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTG 420

Qy      524 AGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGC 583
          ||||| |||| | ||||| || | |||| | || | ||||| || | |||| ||
Db      421 AGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGCGAG 480

Qy      584 AATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCCAT 643
          | | | | | | | | | | |||| | |||| | |||| | |||| | |||| |
Db      481 TCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCAC 540

Qy      644 GTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGC 703
          |||| | || | | ||||| || | | || | |||| | || | ||||| ||
Db      541 GTGGCGGACCAAATGATTGGCAGCTATAATTTGGGGGAATTTCCAGTGGCGAGCGGCGC 600
```

Qy	704	CGGGTCTCCATCGCAGCCCACTGCTCCAGGATCCTAAGGTCATGCTGTTTGTATGAGCCA	763
Db	601	CGAGTTTCCATCGCAGCCCACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCA	660
Qy	764	ACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCT	823
Db	661	ACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCT	720
Qy	824	CGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTTCAGCTC	883
Db	721	CGCAGGGACCGAATTGTGATTGTCACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACAC	780
Qy	884	TTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAA	943
Db	781	TTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCCAGAGGAG	840
Qy	944	ATGCTTGATTCTTCAATGACTGCGGTTACCCCTTGTCCTGAACATTCAAACCCCTTTTGAC	1003
Db	841	ATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTGAT	900
Qy	1004	TTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCC	1063
Db	901	TTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTAC	960
Qy	1064	AAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAAACTTTG	1123
Db	961	AAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTG	1020
Qy	1124	AAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAA	1183
Db	1021	GAGAACATTGAAAGAGCACGATACCTGAAAACCTTACCCATGGTTCCTTTCAAACAAAA	1080
Qy	1184	GATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTG	1243
Db	1081	GATCCTCCTGGGATGTTTCGGCAAGCTTGGTGTCTGCTGAGGCGAGTAACAAGAACTTA	1140
Qy	1244	GTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTT	1303
Db	1141	ATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTTCAAGATCTGATCATGGGCTCTTC	1200
Qy	1304	CTCCTTTTCTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGC	1363
Db	1201	CTCATTTTCTACCTTCTCCGCGTCCAGAACACACGCTAAAGGGCGCTGTGCAGGACCGC	1260
Qy	1364	GTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCGTACACAGGCATGCTGAACGCTGTG	1423
Db	1261	GTGGGGCTGCTCTATCAGCTTGTGGGTGCCACCCCATACACGGGCATGCTCAATGCTGTG	1320
Qy	1424	AATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAG	1483
Db	1321	AATCTGTTTCCCATGCTGAGAGCCGTACGCGACCAGGAGAGTCAGGATGGCCTGTATCAT	1380
Qy	1484	AAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACC	1543
Db	1381	AAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCTCCCCTTCAGCGTCATCGCCACG	1440

XX WO200227016-A2.
XX
PD 04-APR-2002.
XX
PF 25-SEP-2001; 2001WO-US029859.
XX
PR 25-SEP-2000; 2000US-0235268P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.
XX
PI Patel SB, Dean M;
XX
DR WPI; 2002-416483/44.
DR P-PSDB; AAU96986.
XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.
XX
PS Example 3; Page 45-46; 66pp; English.
XX
CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence encodes the rat ABCG5 protein of the invention. (Updated on
CC 07-AUG-2003 to correct OS field.)
XX
SQ Sequence 2035 BP; 481 A; 533 C; 537 G; 484 T; 0 U; 0 Other;

Db	61	CAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTCAGAGGCTCGGCACAG	120
Qy	217	CCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACAT	276
Db	121	CTTAGGTGTCTGAATGTGTCTTCAGCGTCAGCAACCGTGTCTGGGCCCTGGTGGAAACAT	180
Qy	277	CACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGA	336
Db	181	CAAATCATGCCAGCAGAAAGTGGGACAGGAAAATCCTCAAAGATGTCTCCTTGTACATCGA	240
Qy	337	GAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGA	396
Db	241	GAGTGGCCAGACCATGTGCATCTTAGGTAGCTCAGGCTCAGGGAAAACCACGCTGCTGGA	300
Qy	397	CGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGG	456
Db	301	CGCCATCTCTGGGAGGCTGCGGCGCACAGGGACCTTGGAAGGGGAAGTGTGTGTGAACGG	360
Qy	457	CCGGGCGCTGCGCCGGGAGCAGTTCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACAC	516
Db	361	CTGCGAGCTGCGCAGGGACAGTTCGAAGACTGCGTCTCCTACCTCCTGCAGAGCGATGT	420
Qy	517	CCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCG	576
Db	421	CTTCTGAGCAGCCTCACGGTGCGGGAGACGCTGAGATACACGGCGATGCTGGCTCTCCG	480
Qy	577	CCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCT	636
Db	481	CAGCAGCTCCGCGGACTTCTACGACAAGAAGGTAGAGGCAGTCCTGACAGAGCTGAGTCT	540
Qy	637	GAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGA	696
Db	541	GAGCCACGTGGCAGACCAAATGATCGGCAACTATAATTTTGGGGGGATTTCAGTGGCGA	600
Qy	697	GCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGA	756
Db	601	GCGGCGCCGAGTGTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTTGA	660
Qy	757	TGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCTCTGGTGGGA	816
Db	661	CGAGCCAACCACAGGACTGGACTGCATGACTGCAAAATCATATCGTCCTCCTTGGTCTGA	720
Qy	817	ACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTT	876
Db	721	GCTGGCTCGCAGGAACCGCATTGTAATTGTACCATCCACAGCCTCGCTCTGAGCTCTT	780
Qy	877	TCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGGCC	936
Db	781	CCACCATTTCGACAAAATTGCCATTCTGACTTACGGAGAGTTGGTGTCTGTGGCAGGCC	840
Qy	937	AGCGGAAATGCTTGATTTCTTCAATGACTGCGGTTACCCTTGTCTGAACATTCAAACCC	996
Db	841	AGAGGAGATGCTCGGCTTCTTCAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCC	900
Qy	997	TTTTGACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGA	1056

Db 901 CTTTGATTTCTACATGGACTTGACATCGGTGGACACCCAAAGCAGAGAGCGAGAGATAGA 960

Qy 1057 AACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAA 1116
 || | |||| ||||| || | ||||| || | |||| |

Db 961 GACGTACAAGCGAGTCCAGATGCTGGAATCTGCCTTCAGGCAATCGGACATCTGTCACAA 1020

Qy 1117 AACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAA 1176
 || || |||| ||||| || | ||||| |||| | ||||| |||||

Db 1021 AATCCTGGAGAACATTGAAAGAACAAGACACCTGAAAACCCTACCCATGGTTCCTTTCAA 1080

Qy 1177 AACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAG 1236
 ||| ||| || |||| || | |||| || || || ||||| ||||| || ||

Db 1081 AACGAAAATCCTCCCGAATGTTCTGCAAGCTCGGCGTTCTCCTGAGGAGAGTAACGAG 1140

Qy 1237 AAACCTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGG 1296
 |||| | |||| ||||| || | ||||| ||||| ||||| ||||| |||||

Db 1141 AAACCTAATGAGGAATAAGCAGGTGGTGATTATGCGTCTTGTTTCTGAGATCTGATCATGGG 1200

Qy 1297 TTTGTTCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCA 1356
 | ||||| ||||| | |||| || |||| | ||| |||| |||| ||| | ||

Db 1201 TCTGTTCTCATTTTCTACCTTCTCCGAGTCCAGAACACATGCTGAAGGGCGCTGTTCA 1260

Qy 1357 GGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAA 1416
 ||||| ||||| || | ||||| ||||| ||||| ||||| ||||| ||||| ||

Db 1261 GGACCGCGTAGGGCTGTTGTACCAGCTTGTGGGTGCCACCCCGTACACCGGCATGCTCAA 1320

Qy 1417 CGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCT 1476
 ||||| || ||||| |||| | ||||| ||||| ||||| ||||| ||||| |||||

Db 1321 CGCTGTGAACCTCTTTCCCATGCTGAGAGCTGTCAGCGACCAGGAGAGTCAGGATGGCCT 1380

Qy 1477 CTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGT 1536
 ||||| ||||| |||| | ||||| ||||| ||||| | ||||| ||||| | ||

Db 1381 GTACCAGAAGTGGCAGATGCTGCTCGCCTATGTGCTGCATGCTCTCCCCTTCAGCATCGT 1440

Qy 1537 TGCCACCATGATTTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGC 1596
 ||||| ||||| ||||| |||| | ||||| ||||| ||||| |||| | |||| | ||

Db 1441 TGCCACGGTGATTTTTCAGCAGCGTGTGTTACTGGACTCTGGGCTTGTATCCCGAGGTCGC 1500

Qy 1597 CCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAAC 1656
 | ||||| || |||| | |||| | ||||| ||||| ||||| ||||| || ||

Db 1501 CAGATTTGGATACTTCTCTGCCGCTCTGTTGGCCCCCTCACTTAATTGGAGAATTTCTGAC 1560

Qy 1657 TCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCT 1716
 ||||| ||||| ||||| || |||| | ||||| ||||| ||||| |||||

Db 1561 ACTTGTGCTGCTTGGTATGGTCCAAAACCCCAATATTGTCAACAGCATAGTGGCTCTGCT 1620

Qy 1717 GTCCATTGCGGGGGTGCTTGTGGATCTGGATTCTCAGAAACATACAAGAAATGCCCAT 1776
 | ||| | || |||| | ||||| ||||| ||||| ||||| ||||| |||||

Db 1621 GAGTATTTCTGGGTTGCTCATTGGATCTGGATTATCAGAAACATAGAAGAAATGCCCAT 1680

Qy 1777 TCCTTTTAAAATCATCAGTTATTTTACATTCCAAAAATATTGCAGTGAGATTCTTGTAGT 1836
 ||||| ||||| | |||| | |||| | ||||| || || ||||| ||||| ||

Db 1681 TCCTTTTAAAATCCTGGGTTACTTTACCTTCCAAAAGTACTGTTGTGAGATTCTTGTGGT 1740

Qy 1837 CAATGAGTTCTACGGAATTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAATAA 1896
 ||||| ||||| || |||| | ||||| |||| | |||| | |||| | || ||

Db 1741 CAATGAGTTCTATGGCCTGAATTTCACTTGTGGTGGCTCCAACACTTCTGTGCCAATAA 1800

DR P-PSDB; AAU96985.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Example 3; Page 42-43; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the mouse ABCG5 protein of the invention
 XX
 SQ Sequence 1915 BP; 453 A; 502 C; 484 G; 476 T; 0 U; 0 Other;

Query Match 57.1%; Score 1335.8; DB 6; Length 1915;
 Best Local Similarity 81.5%; Pred. No. 0;
 Matches 1560; Conservative 0; Mismatches 352; Indels 3; Gaps 1;

```

Qy      107 ATGGGTGACCTCTCATCTTTGACCCCCGAGGGTCCATGGGTCTCCAAGTAAACAGAGGC 166
      ||||| || | || | || | || | || | || | || | || | || | || | || | || |
Db      1  ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 60

Qy      167 TCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTGGGC 223
      || | ||||| ||||| | | | | || | | | || | || | ||||| | ||
Db      61 TCTCTGAGCTCCCTGGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120

Qy      224 ATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCT 283
      |||| |||| | ||||| ||||| |||| | || | |||| | |||| | |||
Db      121 GTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTCGGGCCTTGGTGGAACATCAAATCA 180

Qy      284 TGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGG 343
      |||| |||| |||| | |||| | ||||| ||||| ||||| ||||| || | |||| | ||
Db      181 TGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGC 240

Qy      344 CAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATG 403
      |||| | ||||| |||| | ||||| |||| | |||| | |||| | |||| | |||| | |||
Db      241 CAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATC 300

Qy      404 TCCGGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCG 463
      ||||| |||| | |||| | | ||||| |||| | |||| | || | | |

```

Db 301 TCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAG 360

Qy 464 CTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCTG 523
 ||||| |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 361 CTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCTGCAGAGCGACGTTTTTCTG 420

Qy 524 AGCAGCCTCACCCTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGC 583
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 421 AGCAGCCTCACTGTGCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGAGC 480

Qy 584 AATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCAT 643
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 481 TCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCAC 540

Qy 644 GTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGC 703
 ||||| ||||| | ||||| ||||| | | ||||| ||||| || ||||| |||||

Db 541 GTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGC 600

Qy 704 CGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCA 763
 || || ||||| ||||| || || ||||| || ||||| || || ||||| |||||

Db 601 CGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCA 660

Qy 764 ACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCT 823
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 661 ACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCCTTCTCTGGCTGAGCTGGCT 720

Qy 824 CGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTTCAGCTC 883
 ||||| ||||| ||||| || ||||| ||||| || ||||| ||||| || || || ||

Db 721 CGCAGGGACCGAATTGTGATTGTCCATCCACCAGCCTCGCTCTGAGCTCTTCCAACAC 780

Qy 884 TTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAA 943
 || ||||| ||||| ||||| | ||||| || | ||||| ||||| ||||| |||||

Db 781 TTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCAGAGGAG 840

Qy 944 ATGCTTGATTTCTTCAATGACTGCGGTTACCCTTGTCCTGAACATTCAAACCCTTTTGAC 1003
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 841 ATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCCTGAACATTCCAATCCCTTTGAT 900

Qy 1004 TTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCC 1063
 || || ||||| ||||| ||||| ||||| ||||| || ||||| ||||| || ||

Db 901 TTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGAAATAGAAACGTAC 960

Qy 1064 AAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTG 1123
 ||| |||| ||||| | |||| |||| |||| |||| | || | ||| |||| | ||

Db 961 AAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTG 1020

Qy 1124 AAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTTCCTTTCAAACCAAA 1183
 |||| ||||| ||||| | ||||| ||||| | ||||| ||||| ||||| |||||

Db 1021 GAGAACATTGAAAGAGCACGATACCTGAAAACCTTACCCACGGTTCCTTTCAAACAAAA 1080

Qy 1184 GATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTG 1243
 ||| ||||| | ||| || || ||||| || ||||| ||||| ||||| |||||

Db 1081 GATCCTCCTGGGATGTTTCGGCAAGCTTGGTGTCTGCTGAGGCGAGTAACAAGAACTTA 1140

Qy 1244 GTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTC 1303
 |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1141 ATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCAGAATCTGATCATGGGCCTCTTC 1200

Qy	1304	CTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGC	1363
Db	1201	CTCATTTCCTACCTTCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGC	1260
Qy	1364	GTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTG	1423
Db	1261	GTGGGGCTGCTCTATCAGCTTGTGGGTGCCACCCCATACACGGCATGCTCAATGCTGTG	1320
Qy	1424	AATCTGTTTCCCGTGCTGCGAGCTGTGACGACCAGGAGAGTCAGGACGGCCTCTACCAG	1483
Db	1321	AATCTGTTTCCCATGCTGAGAGCCGTCAGCGACCAGGAGAGTCAGGATGGCCTGTATCAT	1380
Qy	1484	AAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACC	1543
Db	1381	AAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCTCCCCTTCAGCGTCATCGCCACG	1440
Qy	1544	ATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATT	1603
Db	1441	GTCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTTGTATCCTGAAGTTGCCAGATT	1500
Qy	1604	GGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTG	1663
Db	1501	GGATATTTCTCTGCTGCTCTTTTGGCCCCCTCACTTAATTGGAGAATTTCTAACACTTGTG	1560
Qy	1664	CTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATT	1723
Db	1561	CTGCTTGGTATAGTCCAAAACCCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATC	1620
Qy	1724	GCGGGGGTGCTTGTGGATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTT	1783
Db	1621	TCTGGGCTGCTTATTGGATCTGGATTATCAGAAACATACAAGAAATGCCCATTCCTTTA	1680
Qy	1784	AAAATCATCAGTTATTTTACATTCCAAAATATTGCAGTGAGATTCTTGTAGTCAATGAG	1843
Db	1681	AAAATCCTGGGTATTTTACATTCCAAAATACTGTTGTGAGATTCTCGTGGTCAATGAG	1740
Qy	1844	TTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAATG	1903
Db	1741	TTTTACGGCCTGAACCTCACTTGTGGTGGATCCAACACCTCTATGCTAAATCACCCGATG	1800
Qy	1904	TGTGCCTTCACTCAAGGAATTCATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGA	1963
Db	1801	TGCGCCATCACCAAGGGTCCAGTTCATCGAGAAAACCTGCCAGGTGCTACATCCAGA	1860
Qy	1964	TTCACAATGAACCTTCTGATTTTGTATTTCATTTATTCCAGCTCTTGTATCCTTAG	2018
Db	1861	TTCACGGCAAACCTCCTCATCTTATATGGGTTTATCCCAGCTCTGGTCATCCTAG	1915

RESULT 10

ADB62671

ID ADB62671 standard; cDNA; 2512 BP.

XX

AC ADB62671;

XX

DT 04-DEC-2003 (first entry)

XX
 DE Human cDNA encoding clone LIVER20030650.
 XX
 KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 KW tissue regeneration; cell regeneration; membrane protein;
 KW signal transduction-related protein; transcription-related protein;
 KW osteoporosis; neurological disease; cancer; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1469..2239
 FT /*tag= a
 FT /product= "Clone LIVER20030650 protein"
 XX
 PN EP1308459-A2.
 XX
 PD 07-MAY-2003.
 XX
 PF 28-MAR-2002; 2002EP-00007401.
 XX
 PR 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI; 2003-450961/43.
 DR P-PSDB; ADB64641.
 XX
 PT New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 XX
 PS Claim 1; Page; 222pp; English.
 XX
 CC The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesising the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes

CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.

XX

SQ Sequence 2512 BP; 543 A; 675 C; 701 G; 593 T; 0 U; 0 Other;

Query Match 50.2%; Score 1174.2; DB 9; Length 2512;
Best Local Similarity 71.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 603; Indels 103; Gaps 9;

```
Qy      1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
          |||
Db      81 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 140

Qy     61 GGGTCCG GCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
          |||
Db    141 GGGTCCG GCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 200

Qy    121 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
          |||
Db    201 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 260

Qy    181 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240
          |||
Db    261 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 320

Qy    241 CAGCGTCAGCCACCGC-----GTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGT 296
          |||
Db    321 CAGCGTCAGGTAAGGCAGAGCCCTTGCTGCTGCTGCTCCCCAGGAGTGCGGGGCCCCGGC 380

Qy    297 GGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCA 356
          |||
Db    381 GCTCACCCCTCTGCTGCCTTTCTTCACTCTTTAAGTGCCAGTCTGGGCACTTCGGGCTCC 440

Qy    357 TCCTAGGAAGCTCAGGCTCC-----GGGAAAACCACGCTGCTGGACGCCATGTCC 406
          |||
Db    441 CTCTTTAGTGGATCGGGTGGAGAGAGGAGAGGGAGAAGGGCTGTGCTGGGAAACATGGAG 500

Qy    407 GGGAGGCTGGGGCGCGCGGGGACCTTCTGTTGGGAGGTGTATGTGAACGGCCGGGCGCTG 466
          |||
Db    501 CGACAGTGAATGGCCCTCCCCCTGCCAGGGAAGGGCCTGGGCATAAACAAAGTGGCAG 560

Qy    467 CGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGC 526
          |||
Db    561 CAGTGCCCTGCCAACCCAGTGTCTACGGCCTGCCCTCTGTGGATGGGAATGGGGGTACTG 620

Qy    527 AGCCT-----CACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCAT 573
          |||
Db    621 CGAATGCAAGGAGTCTTGAAACCTGGTGAAAGAATGCAGGGACAGCCACCTCGCAGCCAA 680

Qy    574 CCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAG 633
          |||
Db    681 ACGGACAGGACATTTCAGAGCAACTCCAGCACAGGCCCCCTCCCTACGTGGCAGACAGCCT 740
```

Qy 634 TCTGAGCCATGTGGCAG-----ACCGACTGATTGGCAACTACAGCT 674
 || || || || | | ||||| | |||
 Db 741 CAGTCGCTATCTGCCAGGTTCTACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCC 800

Qy 675 TGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCAT-----CG 716
 || || | | |||| | | | |||
 Db 801 TGTCCGGAGACTACTGGGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGC 860

Qy 717 CAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGG 776
 | | | ||| | | | | |||| | | || | |
 Db 861 CCCTTCCAGGGCCCCAAGCTGCCTTTGCCTTCCTGGGGTTTCTTTAAAGCCACCGCGTG 920

Qy 777 ACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAA 836
 | | || | |||| | | | |||| | || | ||
 Db 921 AGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGAT 980

Qy 837 TTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTTCAG-----CTCTTTGAC 889
 | | || | | | | | | |||| | | |||
 Db 981 GTCTCCTTGACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGG 1040

Qy 890 AAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTT 949
 |||| | | || | | | || | | || | ||
 Db 1041 AAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCTGGGG 1100

Qy 950 GATTTCTTCAATGACTGCGGTTACCCTTGTCCTGAACATTCAAACCTTTTGACTTCTAT 1009
 || | | || || | | | || || | || |||||
 Db 1101 GAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGT---TCCAGGACTGCTTCTCC 1157

Qy 1010 ATGGACCTGACGTGAGTGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGA 1069
 | |||| | | | |||| | |||| |
 Db 1158 TACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTAC 1217

Qy 1070 GTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAAT 1129
 | | |||| | | | |||| | |||| | ||||
 Db 1218 ACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCC 1277

Qy 1130 AT-----TGAAAGAATGAAACACCTGAAAACGTTACCAA 1163
 | | |||| | | |||| |
 Db 1278 GTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTG 1337

Qy 1164 TGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGA 1223
 || | |||| | || | |||| | | | ||||
 Db 1338 GGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCAGCTGCTCCAGGAT 1397

Qy 1224 ---GGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTC 1280
 ||||||||||||||||||||||||||||||||||||
 Db 1398 CCTAGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTC 1457

Qy 1281 AGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGC 1340
 ||||||||||||||||||||||||||||||||||||
 Db 1458 AGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGC 1517

Qy 1341 TAAAGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGT 1400
 ||||||||||||||||||||||||||||||||||||
 Db 1518 TAAAGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGT 1577

Qy 1401 ACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGG 1460

Db	1578	 ACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGG	1637
Qy	1461	AGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCC	1520
Db	1638	 AGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCC	1697
Qy	1521	TCCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCT	1580
Db	1698	 TCCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCT	1757
Qy	1581	TACATCCTGAGGTTGCCCCGATTGGATATTTTTCTGCTGCTCTCTTGGCCCCCACTTAA	1640
Db	1758	 TACATCCTGAGGTTGCCCCGATTGGATATTTTTCTGCTGCTCTCTTGGCCCCCACTTAA	1817
Qy	1641	TTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACA	1700
Db	1818	 TTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACA	1877
Qy	1701	GTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCTCAGAAACA	1760
Db	1878	 GTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATTCTCAGAAACA	1937
Qy	1761	TACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAATATTGCA	1820
Db	1938	 TACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAATATTGCA	1997
Qy	1821	GTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATG	1880
Db	1998	 GTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATG	2057
Qy	1881	TTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATCAATTCAATTGAGAAAA	1940
Db	2058	 TTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATCAATTCAATTGAGAAAA	2117
Qy	1941	CCTGCCCAGGTGCAACATCTAGATTACAAATGAACTTTCTGATTTTGTATTCAATTTATTC	2000
Db	2118	 CCTGCCCAGGTGCAACATCTAGATTACAAATGAACTTTCTGATTTTGTATTCAATTTATTC	2177
Qy	2001	CAGCTCTTGTCACTCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGT	2060
Db	2178	 CAGCTCTTGTCACTCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGT	2237
Qy	2061	AGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGA	2120
Db	2238	 AGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGA	2297
Qy	2121	ACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTCAAGTCTTTTAAC	2180
Db	2298	 ACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTCAAGTCTTTTAAC	2357
Qy	2181	CATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGT	2240
Db	2358	 CATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGT	2417
Qy	2241	TTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGTTATTTGGAAATTTGACTGA	2300

Db 2418 TTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGTTATTTGGAAATTGTGACTGA 2477

Qy 2301 GCGGACCCAAGAATGTAAATAATATTCATAAACCT 2335
 |||

Db 2478 GCGGACCCAAGAATGTAAATAATATTCATAAACCT 2512

RESULT 11

ABK51687

ID ABK51687 standard; cDNA; 1069 BP.

XX

AC ABK51687;

XX

DT 07-AUG-2003 (revised)

DT 30-JUL-2002 (first entry)

XX

DE cDNA encoding hamster ABCG5 protein.

XX

KW Hamster; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ss.

XX

OS Cricetinae.

XX

FH Key Location/Qualifiers

FT CDS 30. .1049

FT /*tag= a

FT /partial

FT /product= "Hamster ABCG5 protein"

FT /note= "This sequence lacks both a start and stop codon"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR P-PSDB; AAU96987.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 47; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart

CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the hamster ABCG5 protein of the invention.
 CC (Updated on 07-AUG-2003 to correct OS field.)

XX

SQ Sequence 1069 BP; 266 A; 282 C; 273 G; 248 T; 0 U; 0 Other;

Query Match 32.9%; Score 770.2; DB 6; Length 1069;
 Best Local Similarity 83.7%; Pred. No. 6.4e-203;
 Matches 896; Conservative 0; Mismatches 173; Indels 2; Gaps 2;

Qy 368 TCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGG 427
 ||||| ||||||||| ||||| ||||| ||||||||| | ||| | |||
 Db 1 TCAGGCTCAGGGAAAACCACGTTGCTGG-TGCCATCTCCGGGAGGCTGCGACGCACAGGG 59

Qy 428 ACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGAC 487
 ||| | ||||||||| ||||||||| | ||||||| ||| ||||||| |||
 Db 60 ACCCTGGAAGGGGAGGTGTTTGTGAACGGCCGTGAGCTGCGCAGGGACCAGTTCCAAGAC 119

Qy 488 TGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACG 547
 ||||||||| ||||||||| ||| | ||||||| ||||| ||||| |||||
 Db 120 TGCTTCTCCTATGTCCTGCAGAGCGACGTCTTTCTGAGCAGTCTCACGGTGCGAGAGACG 179

Qy 548 CTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAG 607
 |||| ||||| ||| ||||||||| ||||| | || | | | ||| |||||
 Db 180 CTGCGCTACACGGCGATGCTGGCCCTCCGCACTAGCTCTTCGGACTTCTATGACAAGAAG 239

Qy 608 GTGGAGGCCGTGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAAC 667
 || ||||| ||||||| ||||||| ||||||||| ||||||||| |||||||||
 Db 240 GTAGAGGCAGTCATGGAAGAGCTAAGTCTGAGCCACGTGGCAGACCGAATGATTGGCAAC 299

Qy 668 TACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTG 727
 || | || ||||| ||||||| || ||||||||| ||||||||| |||||
 Db 300 TATAATTTTGGGGGAATTTCCAGTGCGGAGCGGCGCCGAGTCTCCATCGCAGCCCAACTC 359

Qy 728 CTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACT 787
 | ||||| || ||| ||||| ||||||||| ||||||||| ||||||||| |||||
 Db 360 ATTCAAGACCCCAAGATCATGATGTTTGATGAGCCAACCACAGGACTGGACTGCATGACT 419

Qy 788 GCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTC 847
 || ||||| ||||| ||||||||| || ||||||||| ||||| ||||| | |||
 Db 420 GCAAATCAAATTGTCATCCTCCTGGCAGAGCTGGCTCGCAGGGACCGCATTGTGATCGTC 479

XX
 PN WO200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001WO-US012758.
 XX
 PR 18-APR-2000; 2000US-0198465P.
 PR 15-MAY-2000; 2000US-0204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 XX
 DR WPI; 2002-017598/02.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
 PT useful for screening a compound that increases the level of expression or
 PT activity of SSG polypeptide for treating sterol-related disorder.
 XX
 PS Claim 73; Fig 14B; 105pp; English.
 XX
 CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
 CC identifying a compound useful in the treatment or prevention of a sterol-
 CC related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations, and
 CC for treating any of a large number of SSG associated diseases. The
 CC present sequence is an exon of human SSG DNA
 XX
 SQ Sequence 472 BP; 134 A; 93 C; 100 G; 145 T; 0 U; 0 Other;

Query Match 20.2%; Score 472; DB 6; Length 472;
 Best Local Similarity 100.0%; Pred. No. 2.7e-120;
 Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1869 GCAGCTCAAATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAAT 1928
 |||
 Db 1 GCAGCTCAAATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAAT 60

 Qy 1929 TCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCTGATTTTGT 1988
 |||
 Db 61 TCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCTGATTTTGT 120

 Qy 1989 ATTCATTTATTCCAGCTCTTGTGATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATC 2048
 |||
 Db 121 ATTCATTTATTCCAGCTCTTGTGATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATC 180

Qy 2049 TCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCA 2108
 |||
 Db 181 TCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCA 240
 Qy 2109 TGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTC 2168
 |||
 Db 241 TGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTC 300
 Qy 2169 AAGTCTTTTAACCATTAAGACTCCATTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGC 2228
 |||
 Db 301 AAGTCTTTTAACCATTAAGACTCCATTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGC 360
 Qy 2229 AATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGTTATTTGGA 2288
 |||
 Db 361 AATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGTTATTTGGA 420
 Qy 2289 AATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
 |||
 Db 421 AATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 472

RESULT 13

AAC76065

ID AAC76065 standard; cDNA; 432 BP.

XX

AC AAC76065;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF1620 polynucleotide sequence SEQ ID NO:3239.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerable; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US008621.

XX

PR 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

XX

XX

XX

DR

XX

PT

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

cc
aacc
aaCC
CCcc
cccc
cc

CC

CC

CC

CC

YY

Best Local Similarity 99.8%; Pred. No. 1.7e-108;

|||||

|||||

|||||

JOURNAL OF DOCUMENTATION

Qy 1415 AACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGC 1474
 |||
 Db 241 AACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGC 300
 Qy 1475 CTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTT 1534
 |||
 Db 301 CTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTT 360
 Qy 1535 GTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTT 1594
 |||
 Db 361 GTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTT 420
 Qy 1595 GCCCCGATTTGG 1605
 |||
 Db 421 GCCCCGATTGGG 431

RESULT 14

AAZ94755

ID AAZ94755 standard; cDNA; 281 BP.

XX

AC AAZ94755;

XX

DT 01-AUG-2000 (first entry)

XX

DE Human ATP binding cassette cDNA fragment 168043.

XX

KW ATP binding cassette; human; cholesterol; lipid disorder;

KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;

KW lupus erythematosus; diagnosis; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN WO200018912-A2.

XX

PD 06-APR-2000.

XX

PF 21-SEP-1999; 99WO-EP006991.

XX

PR 25-SEP-1998; 98US-0101706P.

XX

PA (FARB) BAYER AG.

XX

PI Schmitz G, Klucken J;

XX

DR WPI; 2000-293151/25.

XX

PT Adenosine triphosphate binding proteins useful for identifying agents for
 PT treating atherosclerosis and other inflammatory disorders.

XX

PS Claim 9; Page 135; 154pp; English.

XX

CC The present sequence is that of human ATP binding cassette (ABC) cDNA
 CC fragment 168043, identified as a cholesterol-sensitive gene fragment. The
 CC invention provides cholesterol-sensitive ABC genes (see AAZ94734-63).
 CC These genes, and polypeptides encoded by them, can be used for diagnostic

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 38-41; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the upstream genomic sequence, exon 1, intron 1
CC and exon 2 of the human ABCG5 gene located on chromosome 2p21

XX

SQ Sequence 5460 BP; 1351 A; 1350 C; 1508 G; 1243 T; 0 U; 8 Other;

Query Match 10.7%; Score 249.6; DB 6; Length 5460;

Best Local Similarity 98.4%; Pred. No. 7.2e-58;

Matches 252; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4504 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 4563

Qy      61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4564 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 4623

Qy      121 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4624 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 4683

Qy      181 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4684 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 4743
```

Qy 241 CAGCGTCAGCCACCGC 256
 | | | | | | | | | | | |
Db 4744 CAGCGTCAGGTAAGGC 4759

Search completed: February 26, 2004, 01:19:55
Job time : 599.658 secs

OM nucleic - nucleic search, using sw model

Run on: February 26, 2004, 00:48:03 ; Search time 113.204 Seconds
(without alignments)
11471.161 Million cell updates/sec

Title: US-09-989-981A-5
Perfect score: 2340
Sequence: 1 gtcaggtggagcaggcaggg.....aatattcataaacctatggg 2340

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	98	4.2	2418	4	US-09-245-808-2	Sequence 2, Appli
2	96.6	4.1	235	3	US-09-172-108-8	Sequence 8, Appli
3	83	3.5	3376	4	US-09-620-312D-918	Sequence 918, App
4	73.2	3.1	4159	4	US-09-614-912-139	Sequence 139, App
5	61.4	2.6	1551	4	US-09-489-039A-2869	Sequence 2869, Ap
6	61.2	2.6	1722	4	US-09-489-039A-4920	Sequence 4920, Ap
7	59.4	2.5	765	4	US-09-489-039A-3218	Sequence 3218, Ap
8	57	2.4	1668	4	US-09-252-991A-13705	Sequence 13705, A
9	56.8	2.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
10	56.8	2.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
11	56.2	2.4	972	4	US-09-252-991A-9848	Sequence 9848, Ap

	12	56.2	2.4	1713	4	US-09-252-991A-9760	Sequence 9760, Ap
c	13	56.2	2.4	2805	4	US-09-252-991A-10208	Sequence 10208, A
c	14	55.6	2.4	1509	4	US-09-252-991A-13436	Sequence 13436, A
	15	54.2	2.3	840	4	US-09-252-991A-15851	Sequence 15851, A
	16	54	2.3	2175	4	US-09-252-991A-13657	Sequence 13657, A
	17	54	2.3	2292	4	US-09-252-991A-13575	Sequence 13575, A
c	18	53.4	2.3	303	4	US-09-489-039A-4877	Sequence 4877, Ap
	19	53.4	2.3	798	4	US-09-489-039A-4894	Sequence 4894, Ap
	20	53.4	2.3	993	4	US-09-252-991A-10998	Sequence 10998, A
	21	53.4	2.3	1335	4	US-09-252-991A-10934	Sequence 10934, A
c	22	53.4	2.3	2178	4	US-09-252-991A-11254	Sequence 11254, A
c	23	53	2.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	24	53	2.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	25	52.6	2.2	297	4	US-09-489-039A-6200	Sequence 6200, Ap
c	26	52.6	2.2	1800	4	US-09-489-039A-5597	Sequence 5597, Ap
	27	52.6	2.2	1941	4	US-09-489-039A-5579	Sequence 5579, Ap
c	28	52	2.2	474	4	US-09-489-039A-6117	Sequence 6117, Ap
	29	52	2.2	774	4	US-09-489-039A-5958	Sequence 5958, Ap
c	30	51.6	2.2	765	4	US-09-252-991A-3138	Sequence 3138, Ap
	31	51.6	2.2	1476	4	US-09-252-991A-2825	Sequence 2825, Ap
	32	51	2.2	945	4	US-09-252-991A-8027	Sequence 8027, Ap
c	33	51	2.2	1173	4	US-09-252-991A-8287	Sequence 8287, Ap
	34	51	2.2	1209	4	US-09-489-039A-6373	Sequence 6373, Ap
	35	50.6	2.2	765	4	US-09-252-991A-16348	Sequence 16348, A
	36	50.6	2.2	846	4	US-09-252-991A-16477	Sequence 16477, A
c	37	50.6	2.2	855	4	US-09-252-991A-16015	Sequence 16015, A
c	38	50.6	2.2	867	4	US-09-252-991A-15936	Sequence 15936, A
	39	50.6	2.2	2427	4	US-09-252-991A-16255	Sequence 16255, A
	40	50.2	2.1	861	4	US-09-540-236-1011	Sequence 1011, Ap
	41	50.2	2.1	269223	4	US-09-596-002-41	Sequence 41, Appl
	42	50	2.1	744	4	US-09-252-991A-13301	Sequence 13301, A
	43	50	2.1	783	4	US-09-252-991A-12981	Sequence 12981, A
c	44	50	2.1	786	4	US-09-252-991A-12498	Sequence 12498, A
	45	50	2.1	987	4	US-09-489-039A-2741	Sequence 2741, Ap

ALIGNMENTS

RESULT 1

US-09-245-808-2

; Sequence 2, Application US/09245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; TITLE OF INVENTION: encodes it

; FILE REFERENCE: Ross UMb conversion

; CURRENT APPLICATION NUMBER: US/09/245,808

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-2

Query Match 4.2%; Score 98; DB 4; Length 2418;
Best Local Similarity 50.2%; Pred. No. 4e-17;
Matches 242; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

```
Qy      617 GTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTG 676
        |||||  |||| |  |||||  | ||||| |||||  | |||| |  |  |
Db      728 GTCATTCAAGAGTTAGGTCTGGATAAAGTGGCAGACTCCAAGGTTGGAAGTCAAGTTTATC 787

Qy      677 GGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGAT 736
        | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      788 CGTGGTGTGTCTGGAGGAGAAAGAAAAAGGACTAGTATAGGAATGGAGCTTATCACTGAT 847

Qy      737 CCTAAGGTCATGCTGTTTGTATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAG 796
        |||  | | | | | | ||||| || | | | | | | | | | | | |
Db      848 CCTCCATCTTGTCTTGGATGAGCCTACAAGTGGCTTAGACTCAAGCACAGCAAATGCT 907

Qy      797 ATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTAC 856
        | | | | |||||  || | | | | | | | | | | | | | | | |
Db      908 GTCCTTTTGCTCCTGAAAAGGATGTCTAAGCAGGGACGAACAATCATCTTCTCCATTCAT 967

Qy      857 CAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAG 916
        ||||| || | | | | | | | | | | | | | | | | | | | |
Db      968 CAGCCTCGATATTCCATCTTCAAGTTGTTTGTATAGCCTCACCTTATTGGCCTCAGGAAGA 1027

Qy      917 CTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAATGACTGCGGTTACCCT 976
        || | | |||  || | | | | | | | | | | | | | | | |
Db     1028 CTTATGTTCCACGGGCCTGCTCAGGAGGCCTTGGGATACTTTGAATCAGCTGGTTATCAC 1087

Qy      977 TGTCTGAACATTCAAACCCCTTTTGACTTCTATATGGACCTGACGTGAGTGATACCCAA 1036
        |||  |  | | | |||||  ||||| ||||| | | | | | | |
Db     1088 TGTGAGGCCTATAATAACCCCTGCAGACTTCTTCTTGGACATCATTAATGGAGATTCCACT 1147

Qy     1037 AGCAAGGAACGGGAAATAGAAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAG 1096
        || | | | | |||||  | | | | | | | | | | | | | |
Db     1148 GCTGTGGCATTAAACAGAGAAGAAGACTTTAAAGCCACAGAGATCATAGAGCCTTCCAAG 1207

Qy     1097 AA 1098
        |
Db     1208 CA 1209
```

RESULT 2

US-09-172-108-8

; Sequence 8, Application US/09172108

; Patent No. 6160104

; GENERAL INFORMATION:

; APPLICANT: Cunnigham, Mary Jane

; APPLICANT: Zweiger, Gary B.

; APPLICANT: Panzer, Scott R.

; APPLICANT: Seilhamer, Jeffrey J.

; TITLE OF INVENTION: MARKERS FOR PEROXISOMAL PROLIFERATORS

; FILE REFERENCE: PA-0012 US

; CURRENT APPLICATION NUMBER: US/09/172,108
 ; CURRENT FILING DATE: 1998-10-13
 ; NUMBER OF SEQ ID NOS: 56
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 8
 ; LENGTH: 235
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE: -
 ; OTHER INFORMATION: 700138117H1
 US-09-172-108-8

Query Match 4.1%; Score 96.6; DB 3; Length 235;
 Best Local Similarity 68.9%; Pred. No. 2.5e-17;
 Matches 162; Conservative 0; Mismatches 69; Indels 4; Gaps 2;

```

Qy      75 GAAAATTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCCCCG 134
      || ||| | || ||||| || | ||||| |||| | | | ||| || |
Db      1 GAGGATTCACTCACATTTGCTTCCCGCTGGCCATGAGTGAGCTGCCCTTCTGAGTCCAG 60

Qy     135 GAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTG 194
      || ||| || | || ||||| || ||||| ||||| ||| | | |
Db     61 AGGGAGCCAGAGGGCCTCACAAACAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAG 120

Qy     195 CCACCGCCCCGGAGCCT---CACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGCC 251
      || | | | ||| || ||||| | || |||| ||| |||| |||||
Db     121 TTACAGGCTCAGAGGCTCGGCACAGCTTAGGTGTCCTGAATGTGTCCTTCAGCGTCAG-A 179

Qy     252 ACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCA 306
      |||| || ||||| ||||| ||||| ||| |||| |||| |||| |
Db     180 ACCGTGTCGGGCCCTGGTGAACATCAAATCATGCCAGCAGAAGTGGGACAGGAA 234
  
```

RESULT 3

US-09-620-312D-918

; Sequence 918, Application US/09620312D
 ; Patent No. 6569662
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yungqing
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John Tillinghast
 ; APPLICANT: Drmanac, Radoje T.
 ; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
 ; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 918
; LENGTH: 3376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2808)
US-09-620-312D-918

Query Match 3.5%; Score 83; DB 4; Length 3376;
Best Local Similarity 49.6%; Pred. No. 7.9e-13;
Matches 289; Conservative 0; Mismatches 270; Indels 24; Gaps 2;

```
Qy      309 TCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCT 368
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      68 TTCTCAAGTGCCTCTCAGGTAAATTCTGCCGCCGGGAGCTGATTGGCATCATGGGCCCT 127

Qy      369 CAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGA 428
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      128 CAGGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGGAGTCTGGAATGA 187

Qy      429 CCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACT 488
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      188 -----AGGGGCAGATCCTGGTTAATGGAAGGCCACGGGAGCTGAGGACCTTCCGCAAGA 241

Qy      489 GCTTCTCCTACGTCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGC 548
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      242 TGTCTGTCTACATCATGCAAGATGACATGCTGCTGCCGCACCTCACGGTGTTGGAAGCCA 301

Qy      549 TGCCTACACCGCGCTGCTGGCCATCCGCCCGGCAATCCCGGCTCCTTCCAGAAGAAGG 608
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      302 TGATGGTCTCTGCTAACCTGAAGCTGAGTGAGA-----AGCAGGAGG 343

Qy      609 TGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACT 668
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      344 TGAAGAAGGAGCTGGTGACAGAGATCCTGACGGCACTGGGCCTGATGTCGTGCTCCCACA 403

Qy      669 ACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGC 728
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      404 CGAGGACAGCCCTGCTCTCTGGCGGGCAGAGGAAGCGTCTGGCCATCGCCCTGGAGCTGG 463

Qy      729 TCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACACAGGCCTGGACTGCATGACTG 788
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      464 TCAACAACCCGCCTGTCTGTTCTTTGATGAGCCCACAGTGGTCTGGATAGCGCTCTT 523

Qy      789 CTAATCAGATTGTCGTCCTCCTGGTGGAATGGCTCGCAGGAACCGAATTGTGGTTCTCA 848
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      524 GTTTCGAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGGGGGGCGGTACCATCATCTGCA 583
```

Qy 849 CCATTCAACCAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAA 891
 |||| ||||| || | ||| || | |||||
 Db 584 CCATCCACCAGCCCAGTGCCAAGCTCTTTGAGATGTTTGACAA 626

RESULT 4

US-09-614-912-139

; Sequence 139, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

; APPLICANT: Lee, Jian Ming

; APPLICANT: Sakai, Hajime

; APPLICANT: Weng, Zude

; APPLICANT: Caimi, Perry G

; APPLICANT: Anderson, Shawn

; TITLE OF INVENTION: Plant Metabolism Genes

; FILE REFERENCE: BB1378 US NA

; CURRENT APPLICATION NUMBER: US/09/614,912

; CURRENT FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: 60/143,401

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/143,412

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/146,650

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/170,906

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: 60/172,959

; PRIOR FILING DATE: 1999-12-21

; PRIOR APPLICATION NUMBER: 60/172,946

; PRIOR FILING DATE: 1999-12-21

; NUMBER OF SEQ ID NOS: 204

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 139

; LENGTH: 4159

; TYPE: DNA

; ORGANISM: Oryza sativa

US-09-614-912-139

Query Match 3.1%; Score 73.2; DB 4; Length 4159;

Best Local Similarity 50.5%; Pred. No. 5.1e-10;

Matches 205; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

Qy 647 GCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCCCGG 706
 || ||| | | ||| || | | | |||| || ||| ||| || | ||
 Db 414 GCGGACACGATCGTCGGCGACCAGATGCAGAGGGGGATCTCCGGTGGTCAGAAGAAACGC 473
 Qy 707 GTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACC 766
 ||| ||| || | | | || | |||| | | | ||||| | |
 Db 474 GTCACCACCGGTGAGATGATTGTCGGTCCAACAAAGGTTCTATTCATGGATGAGATATCA 533

Qy 767 ACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGC 826
 || || |||| || || | ||||| || | || | | |
 Db 534 ACTGGATTGGACAGCTCCACCACATTCCAGATTGTCAAATGCCTTCAGCAAATCGTGCAC 593

Qy 827 AGGAACCGAATTGTGGTTCTCA---CCATTACCAGCCCCGTTCTGAGCTTTTTCAGCTC 883
 | | | |||| | | | || | |||| |||| ||||
 Db 594 TTGGGCGAGGCAACCATCCTCATGTCACTCCTACAACCAGCCCCTGAGACTTTTGAGCTA 653

Qy 884 TTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAA 943
 || || | || | | || | | | || | | | | |
 Db 654 TTCGATGACATTATCCTACTGTGAGAAGGCCAGATTGTTTATCAGGGACCCCGGAATAC 713

Qy 944 ATGCTTGATTCTTCAATGACTGCGGTTACCTTGTCCTGAACATTCAAACCCTTTTGAC 1003
 | |||| |||| | |||| | || | || | | | |
 Db 714 GTCCTTGAGTTCTTTGAGTCATGCGGATTCCGCTGCCAGAGCGTAAGGGTACTGCAGAC 773

Qy 1004 TTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGG 1049
 || | || | || | || | | || | || |
 Db 774 TTTCTTCAGGAGGTGACATCAAAGAAGGATCAGGAGCAGTATTGGG 819

RESULT 5

US-09-489-039A-2869
 ; Sequence 2869, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 2869
 ; LENGTH: 1551
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-2869

Query Match 2.6%; Score 61.4; DB 4; Length 1551;
 Best Local Similarity 48.3%; Pred. No. 5.8e-07;
 Matches 272; Conservative 0; Mismatches 276; Indels 15; Gaps 3;

Qy 306 AGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAA 365
 || | || |||| || | ||||| |||| | | |||| | ||
 Db 851 AGGTGCTGAAAGGCATCGATCTGCAGGTGGAGAACGGGGAGGTGATCAGCATCATCGGCC 910

Qy 366 GCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGG-----AGGCTGGGGC 419
 || ||||| ||||| || | | |||| || || |
 Db 911 CGTCCGGCTCCGGCAAAACACCTGATCCGCACCATCAACGCCCTCGAAAGCCTTGATG 970

Qy 420 GCGCGGGGACCTTCCTG---GGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGC 476
 ||| || | | | || || |||| || | || || | | | |
 Db 971 GCGGGGAGATCATTCTCTACGGCGAGGACTATCTTAAGGGCGGAGCCATCGTCGACAAAC 1030

Qy 477 AGTTCCAGGACTGCTTCTCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCG 536
 | | | | | | | | | | | | | | |
 Db 1031 GCCAGATGCGCGCCGGGTACGGCGCATCGGCATGGTCTTCCAGAGCTTCAACCTGTTCC 1090

Qy 537 TGC GCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCT 596
 | | | | | | | | | | | | | | | | | | |
 Db 1091 CCCACCGCACGGTGCTCGACAACGTGATGCTGGCCCCGC-----GCTATCACCAGCTGC 1144

Qy 597 TCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGAC 656
 | | | | | | | | | | | | | | | | | | |
 Db 1145 TGGACCAGCCGGTCGCCCCGAGCAGGCCCTGGCGCTGCTCGACCGCGTCGGCCTGCTGG 1204

Qy 657 TGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCG 716
 | | | | | | | | | | | | | | | | | | |
 Db 1205 CCCATGCCACAAGTACCCCGGACAGCTCTCCGGCGGCCAGCAGCAGCGCTGGCGATCG 1264

Qy 717 CAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGG 776
 | | | | | | | | | | | | | | | | | | |
 Db 1265 CCCGGGCGCTGGCGCTGAAGCCGGACATTATGCTGTTTGACGAACCGACCTCGGCGCTGG 1324

Qy 777 ACTGCATGACTGCTAATCAGATTGTTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAA 836
 | | | | | | | | | | | | | | | | | | |
 Db 1325 ATCCGGAGCTGGTAGGCGAAGTGCTGAAGGTCATTAGTCGCTGGCCCGCGAAGGCATGA 1384

Qy 837 TTGTGGTTCTCACCATTACCCAG 859
 | | | | | | | | | | | | | | | | | | |
 Db 1385 CCATGCTGATTGTCACTCACGAG 1407

RESULT 6

US-09-489-039A-4920

; Sequence 4920, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 4920

; LENGTH: 1722

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-4920

Query Match 2.6%; Score 61.2; DB 4; Length 1722;

Best Local Similarity 46.4%; Pred. No. 7e-07;

Matches 284; Conservative 0; Mismatches 313; Indels 15; Gaps 2;

Qy 270 GGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGT 329
 | | | | | | | | | | | | | | | | | | |

```

Db      1025 GGGAGGTCACCTTCCGCTATCCTCAGCAGCCCTCCCCTGCCCTGGAGAATATTTCCCTGC 1084
QY      330 ACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGC 389
        | | |   |||  | |  |||  |||  | |  | |  |||  |||  |||  | |||
Db      1085 AGATTGCCGCCGGAGAGCACATCGCCATTCTTGCCCGGACCGCTGCGGAAAATCGACGC 1144
QY      390 TGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCTTGGGGGAGGTGTATG 449
        || || |   | ||   |||  |||  |||  |||  |||  |||  | |  | |
Db      1145 TGTTCAGTTGCTTACCC-----GCGCCTGGGACCCGTCACAGGGAGAGATTCTG 1194
QY      450 TGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCTACGTCCTGCAGA 509
        |   |   | | | |   |||   |||  |   |   |   |   |
Db      1195 CTCAACAATCAGCCGCTCTCCGGCCTCAGCGAAGCCACTCTTCGGCAGGC-----AATGA 1249
QY      510 GCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGG 569
        |||   | | | | |   |   |   |||  |||  |||  | |||  | |||
Db      1250 GCGTAGTGCCGCAGCGCGTGCACCTGTTACGCGCCACCTGCGCGACAACCTGCTGCTGG 1309
QY      570 CCATCCGCCCGCGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGC 629
        |   |   |   | | |   ||   ||   |   |||  ||   ||   |
Db      1310 CGGCGCCTGAAGCGGATGACGCTCATCTCAGCGCTACCCTTGAGAAGGTGGGCCTCGAAA 1369
QY      630 TGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTC 689
        |||   |||  |   |   |   |   |   |   |   |   |   |
Db      1370 AACTGCTGCAAGATGGTGGTCTTAACGGCTGGCTGGGCGAAGGCGGGCGTCAGCTCTCCG 1429
QY      690 CGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGC 749
        || || | |||||  | ||||  ||   | ||||| |||  |   | |||
Db      1430 GCGGCGAACTGCGCCGACTGGCCATTGCCCGCGCGCTGCTCCATGATGCGCCGCTGATGC 1489
QY      750 TGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTGCTCCTCC 809
        || | ||||| || ||   |||  ||||   || | | |||||  |   || |
Db      1490 TGCTCGATGAACCGACAGAAGGTCTGGATGCGGCCACCGAAAGCCAGATCCTGCATCTAC 1549
QY      810 TGGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCCAGCCCCGTTCTG 869
        |||  || |   |||  ||   |   || | |||  || |||  |
Db      1550 TGGCAGATGTCATGCGCGACAAAACCGTGCTGATGGTGACCCATCGCCTGCGGGACCTGG 1609
QY      870 AGCTTTTTTCAGC 881
        |   ||||  | |
Db      1610 CGGGTTTTTAATC 1621

```

RESULT 7

US-09-489-039A-3218

; Sequence 3218, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3218
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3218

Query Match 2.5%; Score 59.4; DB 4; Length 765;
Best Local Similarity 45.4%; Pred. No. 1.4e-06;
Matches 262; Conservative 0; Mismatches 306; Indels 9; Gaps 1;

```
Qy      286 CCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCA 345
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      30 CTGGAAGGCAGGCAAAAAGGTCATCGTCAATAATGTCTCGCTGCGGGTGCCGCGAGGCGA 89

Qy      346 GATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTC 405
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db      90 AACGGTCGGAAGTGTGGGGCCCAACGGCTGCGGCAAATCCTCGCTGCTGCGCGTTCTGGC 149

Qy      406 CGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCT 465
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     150 GGGCCTGCGCCGCCCGGATGCAGGTGCGCTCACCTCGACGGCCAGGATATCGCCCGGAT 209

Qy      466 GCGCCGGGAGCAGTTCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCTGAG 525
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     210 GGCGAAAAAGCAGCTCGCCCGCCGCGTGGCTTTCGTCGAGCAACACGGCATGACCGAGGC 269

Qy      526 CAGCCTCACCGTGCGGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAA 585
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     270 CAATATGCGGGTGCGCGACGTCGTGCGC-----CTGGGACGCATTCCCCACCACTC 320

Qy      586 TCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGT 645
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     321 TCCGTTCTCAAACCTGGAGCGCTCAGGATGACGAGGCGATTGCCGCCGCGCTGCAGCGGGT 380

Qy      646 GGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCG 705
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     381 AGCGATGCTGGAGAAAAGCGAACAGGGATGGTTAAGCCTCTCCGGCGGCGAGCGGCAGCG 440

Qy      706 GGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAAC 765
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     441 GGTGCATATCGCCCGCGCGCTGGCGCAGAGCCCGAGCGAAATCCTGCTGGATGAGCCGAC 500

Qy      766 CACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCTCCTCCTGGTGGAAGTGGCTCG 825
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     501 CAACCATCTGGATATACACCATCAGATGCAGTTAATGCAGTTGATCAGCGAGCTGCCGGT 560

Qy      826 CAGGAACCGAATTGTGGTTCTCACCATTACCCAGCCC 862
      | | | | | | | | | | | | | | | | | | | | | | | | |
Db     561 AACCAGCATTGTGGCCATTACGATCTTAACCATGCC 597
```

RESULT 8
US-09-252-991A-13705
; Sequence 13705, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:


```
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13705
; LENGTH: 1668
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13705
```

```
Query Match          2.4%; Score 57; DB 4; Length 1668;
Best Local Similarity 45.7%; Pred. No. 1e-05;
Matches 238; Conservative 0; Mismatches 280; Indels 3; Gaps 1;
```

```
Qy      312 TCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAG 371
      || | | || | | | | | | | | | | | | | | | |
Db      959 TCGACGGGGTCAATTCGAACTACCCCGCGGGCAGACGCTGGGCATCGTTGGCGAAAGCG 1018

Qy      372 GCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCT 431
      |||| || || | || || | || | | | | | | | | |
Db     1019 GCTCGGGCAAGTCGACCCCTTGGCCTGGCAATCCTGCGGCTGCTGGAAAGCCAGGGCGGCA 1078

Qy      432 TCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCAGGACTGCT 491
      ||| | | | | | | | | | | | | | | | | |
Db     1079 TCCGCTTCGAAGGCACCCGGCTGGACGGTCTCGCGCAACATGACGTGCGCCCGCTGCGCC 1138

Qy      492 TCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGC 551
      | | | | | | | | | | | | | | | | | | |
Db     1139 GCGAGATGCAGGTGGTGTTCAGGACCCATATGGCAGCCTCAGCCCACGCATGTGTGTCTG 1198

Qy      552 ACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGG 611
      | | | | | | | | | | | | | | | | | | |
Db     1199 GCGAGATCGTCGGCGAAGGCCTGCGCATCCATAGGATCGGCAGCGAGGCCGAACAGGAGC 1258

Qy      612 AGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACA 671
      |||| |||| | | |||| | | | | | | | | | |
Db     1259 AGGCGATCATCGACGCGCTG---GTGGAGGTGCGGGCTCGATCCGCAGACCCGCTACCGTT 1315

Qy      672 GCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCC 731
      | | | | | | | | | | | | | | | | | | |
Db     1316 ACCCCACGAATTCCTCGGCGGCCAGCGCCAGCGCATCGCCATCGCCCGGGCACTGGTGC 1375

Qy      732 AGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTA 791
      | | | | | | | | | | | | | | | | | | |
Db     1376 TGAAACCGGCACTGATCCTGCTCGACGAACCGACCTCGGCGCTCGACCGCACCGTGCAGC 1435

Qy      792 ATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAAC 832
      ||| | | | | | | | | | | | | |
Db     1436 GCCAGGTCGTGGAATTGCTGCGGCAACTGCAGGGCAAGTAC 1476
```

RESULT 9

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 2.4%; Score 56.8; DB 3; Length 4403765;

Best Local Similarity 46.0%; Pred. No. 0.0012;

Matches 273; Conservative 0; Mismatches 312; Indels 9; Gaps 2;

Qy 271 GGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTA 330

|| || | || | | || || || || | || |

Db 2879474 GGTCGTCGAGTATTCCAGCGGCGGGTACGCCGTGCGGCCGATCGACGGGTTAAGCCTCGA

2879533

Qy 331 CGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCT 390

|||| | || | | || | | || | | || | || | || | || |

Db 2879534 CGTGGCGCCGGGGTCGCTGGTGATCTTGCTTGGGCCAGCGGCTGCGGGAAGACGACCCT

2879593

Qy 391 GCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGACCTTCCTGGGGGAGGTGTATGT 450

|| | | | || | || | | | || | | | || |

Db 2879594 CTTGTCCTGCCTCGGCGGCATCCTGCGCCCGAAGTCCGGCTCAATCAAGTTGACGATGT

2879653

Qy 451 GAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAG 510

|| | | |||| | | | | | || | | ||

Db 2879654 CGACATCACGACGCTGGAGGGCGCCGCGCTGGCGAAGTATCGGCGTGACAAGGTAGGGAT

2879713

Qy 511 CGACACCCTGCTGAGCAGCCTACCGTGCGCGAGACGCTGCACTACCCGCGCTGCTGGC 570

|| | || | | || || | | || | | | || | || |

Db 2879714 CGTCTTCCAGGCGTTCAACCTGGTCTCGAGCCTTACCGCCCTGGAGAACGTGATGGTCCC

2879773

Qy 571 CATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCT 630
 | | | | | | | | | | | | | | | | | | | | | |
 Db 2879774 GCTGCGCGCGGCCGG---CGTGTACAGAGCGGCCGCGCGTAAGCGTGCCGAGGACCTGCT
 2879830

Qy 631 GAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCAC 690
 | | | | | | | | | | | | | | | | | | | | | |
 Db 2879831 GATCCGAGTCAATCTCGCGAACGAATG-----AAACACGCCCCGGGTGACATGAGCGG
 2879884

Qy 691 GGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCT 750
 | | | | | | | | | | | | | | | | | | | | | |
 Db 2879885 CGGCCAGCAGCAACGCGTCGCGGTGCGCCGCGCGATCGCGCTGGACCCGCAATTGATCCT
 2879944

Qy 751 GTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCT 810
 | | | | | | | | | | | | | | | | | | | | | |
 Db 2879945 TGCCGACGAACCGACCGCGCACCTGGACTTCATCCAGGTGGAGGAGGTGCTGCGGCTGAT
 2880004

Qy 811 GGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACACAGCCCCG 864
 | | | | | | | | | | | | | | | | | | | | | |
 Db 2880005 CCGCTCGCTAGCGCAGGGCGACCGTGTGGTGGTGGTCGCGACCCACGACAGCCG 2880058

RESULT 10

US-09-103-840A-1

; Sequence 1, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 4411529
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; OTHER INFORMATION: H37Rv
 US-09-103-840A-1

Query Match 2.4%; Score 56.8; DB 3; Length 4411529;
 Best Local Similarity 46.0%; Pred. No. 0.0012;
 Matches 273; Conservative 0; Mismatches 312; Indels 9; Gaps 2;

Qy 271 GGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTA 330
 | | | | | | | | | | | | | | | | | | | | | |
 Db 2883366 GGTCGTCGAGTATTCCAGCGGCGGGTACGCCGTGCGGCCGATCGACGGGTTAAGCCTCGA
 2883425

Qy 331 CGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCT 390
 |||| | || | | | | | | | | | |||| |||| | | ||
 Db 2883426 CGTGGCGCCGGGGTCTGCTGGTGATCTTGCTTGGGCCAGCGGCTGCGGGAAGACGACCT
 2883485

Qy 391 GCTGGACGCCATGTCCGGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGAGGTGTATGT 450
 || | | | | || | || | | | | | | | | | | ||||
 Db 2883486 CTTGTCTGCTCGGCGGCATCCTGCGCCCGAAGTCCGGCTCAATCAAGTTTGACGATGT
 2883545

Qy 451 GAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCTACGTCCTGCAGAG 510
 || | | |||| | | | | | | | | | | | | ||
 Db 2883546 CGACATCACGACGCTGGAGGGCGCCGCGCTGGCGAAGTATCGGCGTGACAAGGTAGGGAT
 2883605

Qy 511 CGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACGCGCTGCTGGC 570
 || | || | | || || | | || | | | | | | | | |
 Db 2883606 CGTCTTCCAGGCGTTCAACCTGGTCTCGAGCCTTACCGCCCTGGAGAACGTGATGGTCCC
 2883665

Qy 571 CATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTGATGGCAGAGCT 630
 | ||| | | || || | | | | | | | | | | ||||
 Db 2883666 GCTGCGCGCGGCCGG---CGTGTCACGAGCGGCCGCGCGTAAGCGTGCCGAGGACCTGCT
 2883722

Qy 631 GAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCAC 690
 || | | || | | || || || | | || | | || | | |
 Db 2883723 GATCCGAGTCAATCTCGGCGAACGAATG-----AAACACCGCCCGGTGACATGAGCGG
 2883776

Qy 691 GGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCT 750
 || ||| || || || | |||| | | | | |||| | | ||||
 Db 2883777 CGGCCAGCAGCAACGCGTCTGCGGTGCGCCGCGCGATCGCGTGGACCCGCAATTGATCCT
 2883836

Qy 751 GTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCT 810
 || || || ||| | |||| |||| || | | || | | || |
 Db 2883837 TGCCGACGAACCGACCGCGCACCTGGACTTCATCCAGGTGGAGGAGGTGCTGCGGCTGAT
 2883896

Qy 811 GGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCG 864
 || || | | |||| | |||| | | | | || | ||||
 Db 2883897 CCGCTCGCTAGCGCAGGGCGACCGTGTGGTGGTGGTGGCGACCCACGACAGCCG 2883950

RESULT 11

US-09-252-991A-9848

; Sequence 9848, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9848
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9848

Query Match 2.4%; Score 56.2; DB 4; Length 972;
Best Local Similarity 45.6%; Pred. No. 1.3e-05;
Matches 276; Conservative 0; Mismatches 323; Indels 6; Gaps 2;

```
Qy      311 CTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCA 370
          || || | || || | | |||| || | | | | || ||
Db      88 CTGAACGGCGTATCGTTTGAAGTGAAGCCGGCAAGACCCTCGCCGTGGTCGGCGAGTCC 147

Qy      371 GGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACC 430
          |||| || | || || | || || | || || || || || |
Db      148 GGCTGCGGCAAGTCGACCCTGGCGCGCGCCCTGACCCTGATCGAGGAACCCACCTCCGGC 207

Qy      431 TTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGC 490
          | ||| | | || || | || || || || || || || |
Db      208 TCGCTGAAAATCGCCGGGCAGGAGGTCAAGGGCGCCAGCAAGGACCAGCGCCGGCA---G 264

Qy      491 TTCTCCTACGTCTCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTG 550
          || | || | |||| || | || | | | | | | | |
Db      265 TTGCGCCGCGACGTGCAGATGGTCTTCCAGAACCCTACGCCTCGCTCAATCCGCGACAG 324

Qy      551 CACTACACCGCGCTGCTGGCCATCCGCCCGGCAATCCCGGCTCCTTCCAGAAGAAGGTG 610
          | | | || | |||| | | | | | | | | | |||
Db      325 AAGATCGGCGACCAAGTTGGCCGAGCCGCTGCTGATCAACACCGCGCTGTCGCGCGAGGAA 384

Qy      611 GAGGCCGTTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTAC 670
          || | || |||| || || |||| || | || | | || |
Db      385 CGCCGCGAAAAGGTCCAGCAGATGATGCCAGGTGCGCCTGCGGCCGGAGCATTACCAG 444

Qy      671 AGCTTGGGGGGCATTTCACGGGTG---AGCGGCGCCGGGTCTCCATCGCAGCCCAGCTG 727
          ||| | | || | | || | |||| | | || || || || ||
Db      445 CGCTACCCGCACATGTTCTCCGGCGGCCAGCGCCAGCGCATCGCCCTGGCCCGGGCGATG 504

Qy      728 CTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACT 787
          | | | || |||| || || || |||| || | | || || ||
Db      505 ATGCTGCAACCCAAGGTGCTGGTGGCGGACGAGCCGACCTCGGCCCTCGACGTGTCGATC 564

Qy      788 GCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTC 847
          ||| | | || | |||| || | || | | || | | |
Db      565 CAGGCCCAGGTACTGAACCTTTTCATGGACCTGCAGCAGCAGTTCCGCACCGCCTACGTG 624

Qy      848 ACCATTACACGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGC 907
          ||| || || | || | || || || || | | || | |
Db      625 TTCATCTCGCACAACCTGGCGGTGGTCCGCCACGTCGCCGACGACGTCCTGGTGATGTAC 684

Qy      908 TTCGG 912
          ||||
```

```

US-09-252-991A-9760
; Sequence 9760, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9760
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9760

```

Query Match 2.4%; Score 56.2; DB 4; Length 1713;
Best Local Similarity 45.6%; Pred. No. 1.8e-05;
Matches 276; Conservative 0; Mismatches 323; Indels 6; Gaps 2;

Qy	311	CTCAAAGATGTCTCCTTGACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCA	370
Db	62	CTGAACGGCGTATCGTTTGAAGTGAAGCCGGCAAGACCCTCGCCGTGGTCGGCGAGTCC	121
Qy	371	GGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACC	430
Db	122	GGCTGCGGCAAGTCGACCCTGGCGCGCGCCCTGACCCTGATCGAGGAACCCACCTCCGGC	181
Qy	431	TTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGC	490
Db	182	TCGCTGAAAATCGCCGGGCAGGAGGTCAAGGGCGCCAGCAAGGACCAGCGCCGGCA---G	238
Qy	491	TTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGC GCGAGACGCTG	550
Db	239	TTGCGCCGCGACGTGCAGATGGTCTTCCAGAACCCTACGCCTCGTCAATCCGCGACAG	298
Qy	551	CACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTG	610
Db	299	AAGATCGGCGACCAGTTGGCCGAGCCGCTGCTGATCAACACCGCGTGTGCGCGGAGGAA	358
Qy	611	GAGGCCGT CATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTAC	670
Db	359	CGCCGCGAAAAGGTCCAGCAGATGATGCGCCAGGTCGGCCTGCGGCCGGAGCATTACCAG	418
Qy	671	AGCTTGGGGGGCATTTCACGGGTG---AGCGGCGCCGGGTCTCCATCGCAGCCCACTG	727
Db	419	CGCTACCCGCACATGTTCTCCGGCGGCCAGCGCCAGCGCATCGCCCTGGCCCCGGGCGATG	478

Qy 728 CTCCAGGATCCTAAGGTCATGCTGTTGATGAGCCAACCACAGGCCTGGACTGCATGACT 787
 | | | | | | | | | | | | | | | | | | | |
 Db 479 ATGCTGCAACCCAAGGTGCTGGTGGCGGACGAGCCGACCTCGGCCCTCGACGTGTCGATC 538
 Qy 788 GCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTC 847
 | | | | | | | | | | | | | | | | | | | |
 Db 539 CAGGCCCAGGTACTGAACCTTTTCATGGACCTGCAGCAGCAGTTCCGCACCGCCTACGTG 598
 Qy 848 ACCATTACACGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGC 907
 | | | | | | | | | | | | | | | | | | | |
 Db 599 TTCATCTCGCACAACCTGGCGGTGGTCCGCCACGTCGCCGACGACGTCCTGGTGATGTAC 658
 Qy 908 TTCGG 912
 | | | |
 Db 659 CTCGG 663

RESULT 13

US-09-252-991A-10208/c

; Sequence 10208, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 10208

; LENGTH: 2805

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-10208

Query Match 2.4%; Score 56.2; DB 4; Length 2805;

Best Local Similarity 45.6%; Pred. No. 2.4e-05;

Matches 276; Conservative 0; Mismatches 323; Indels 6; Gaps 2;

Qy 311 CTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCA 370
 | | | | | | | | | | | | | | | | | | | |
 Db 1629 CTGAACGGCGTATCGTTTGAAGCCGGCAAGACCCTCGCCGTGGTCGGCGAGTCC 1570
 Qy 371 GGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGGGGGACC 430
 | | | | | | | | | | | | | | | | | | | |
 Db 1569 GGCTGCGGCAAGTCGACCCTGGCGCGCGCCCTGACCCTGATCGAGGAACCCACCTCCGGC 1510
 Qy 431 TTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGC 490
 | | | | | | | | | | | | | | | | | | | |
 Db 1509 TCGCTGAAAATCGCCGGGCAGGAGGTCAAGGGCGCCAGCAAGGACCAGCGCCGGCA---G 1453

```

Qy      491 TTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTG 550
      || | | | | | | | | | | | | | | | | | | | |
Db      1452 TTGCGCCGCGACGTGCAGATGGTCTTCCAGAACCCCTACGCCTCGCTCAATCCGCGACAG 1393

Qy      551 CACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTG 610
      | | | | | | | | | | | | | | | | | | | |
Db      1392 AAGATCGGCGACCAGTTGGCCGAGCCGCTGCTGATCAACACCGCGCTGTGCGCGAGGAA 1333

Qy      611 GAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTAC 670
      || | | | | | | | | | | | | | | | | | |
Db      1332 CGCCGCGAAAAGGTCCAGCAGATGATGCGCCAGGTGCGCCTGCGGCCGGAGCATTACCAG 1273

Qy      671 AGCTTGGGGGGCATTTCCACGGGTG---AGCGGCGCCGGGTCTCCATCGCAGCCCAGCTG 727
      || | | | | | | | | | | | | | | | | | |
Db      1272 CGCTACCCGCACATGTTCTCCGGCGGCCAGCGCCAGCGCATCGCCCTGGCCCGGGCGATG 1213

Qy      728 CTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACT 787
      | | | | | | | | | | | | | | | | | | | |
Db      1212 ATGCTGCAACCCAAGGTGCTGGTGGCGGACGAGCCGACCTCGGCCCTCGACGTGTCGATC 1153

Qy      788 GCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGTTCTC 847
      || | | | | | | | | | | | | | | | | | |
Db      1152 CAGGCCCAGGTACTGAACCTTTTCATGGACCTGCAGCAGCAGTTCCGCACCGCCTACGTG 1093

Qy      848 ACCATTACCCAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCATCCTGAGC 907
      || | | | | | | | | | | | | | | | | | |
Db      1092 TTCATCTCGCACAACTGGCGGTGGTCCGCCACGTCGCCGACGACGTCCTGGTGATGTAC 1033

Qy      908 TTCGG 912
      ||||
Db      1032 CTCGG 1028

```

RESULT 14

US-09-252-991A-13436/c

; Sequence 13436, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13436

; LENGTH: 1509

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13436

Query Match

2.4%; Score 55.6; DB 4; Length 1509;

Best Local Similarity 46.0%; Pred. No. 2.4e-05;
Matches 227; Conservative 0; Mismatches 264; Indels 3; Gaps 1;

```
Qy      339 GCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACG 398
          |||||  || ||||| | || ||||| || || || || || ||
Db      1502 GCGGGCAGACGCTGGGCATCGTTGGCGAAAGCGGCTCGGGCAAGTCGACCCTTGGCCTGG 1443

Qy      399 CCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCC 458
          | || | || ||||| || || | || | || | || | || |
Db      1442 CAATCCTGCGGCTGCTGGAAAGCCAGGGCGGCATCCGCTTCGAAGGCACCCGGCTGGACG 1383

Qy      459 GGGCGTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCC 518
          | | | | || | || | || | || | || | || | || |
Db      1382 GTCTCGCGCAACATGACGTGCGCCCGCTGCGCCGCGAGATGCAGGTGGTGTTCAGGACC 1323

Qy      519 TGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCAC'TACACCGCGCTGCTGGCCATCCGCC 578
          ||||| || | || | | | | | | | | | | | | | |
Db      1322 CATATGGCAGCCTCAGCCACGCATGTGTGTGCGCGAGATCGTCGGCGAAGGCCTGCGCA 1263

Qy      579 GCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGA 638
          | | | |||| | | || | |||| |||| | | |||| |
Db      1262 TCCATAGGATCGGCAGCGAGGCCGAACAGGAGCAGGCGATCATCGACGCGCTGGTGGAGG 1203

Qy      639 GCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGC 698
          | | | || | || | || || || | | | | | | || | || |
Db      1202 TC---GGGCTCGATCCGCAGACCCGCTACCGTTACCCCCACGAATTCTCCGGCGGCCAGC 1146

Qy      699 GGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTGATG 758
          | | || || ||||| || | || | || | || | || | || |
Db      1145 GCCAGCGCATCGCCATCGCCCGGGCACTGGTGCTGAAACCGGCAC'TGATCCTGCTCGACG 1086

Qy      759 AGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCTCCTGGTGGAAAC 818
          | || || | | || || || || || || | || | || | || |
Db      1085 AACCGACCTCGGCGCTCGACCGCACCGTGCAGCGCCAGGTCGTGGAATTGCTGCGGCAAC 1026

Qy      819 TGGCTCGCAGGAAC 832
          || || | || |
Db      1025 TGCAGGGCAAGTAC 1012
```

RESULT 15

US-09-252-991A-15851

; Sequence 15851, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 15851
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15851

Query Match 2.3%; Score 54.2; DB 4; Length 840;
Best Local Similarity 46.4%; Pred. No. 4.3e-05;
Matches 253; Conservative 0; Mismatches 283; Indels 9; Gaps 2;

```
Qy      318 ATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCG 377
      || ||| || | | || | ||| || | || | || |
Db      107 ATCTCTCGCTGGCCATCCCCGAGGGTTCGTTTCAGTGTGATCGTCGGGCCAACGCCTGCG 166

Qy      378 GGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGG 437
      | || | || ||||| || |||| | | ||| || || | | || |
Db      167 GCAAGTCGACCCTGCTGGCGGCATTGTCTGCGCCTGTTGGCGCCGGCCGAGGGCCGGGTGG 226

Qy      438 GGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTTCAGGACTGCTTCTCCT 497
      | || | | || || ||| | | | | | | | | |
Db      227 TGCTGGACGGCAGGGATATCCACAGCCTGCCGGGACGGGAAGTGGCGCGGCGTCTCGGCC 286

Qy      498 ACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCAGGAGACGCTGCACTACA 557
      | | ||||| || | || ||||| ||| || | || |
Db      287 TGCTGCCGCAGAGCGCGCTGGCGCCGGATGGCATCACGGTGGCCGAGCTGGTGGCGCGC- 345

Qy      558 CCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCG 617
      | || ||||| | | | ||| || | | ||| || |
Db      346 -----GGGCGCTATCCGCACCAGTCGTTCTGCGCCAGTG-GTCGCCGGCGGATGAGC 397

Qy      618 TCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGG 677
      | || || | | || | || | || | || | || | || |
Db      398 GCGCGGTAGCGGCGGCGTTACGCGCCACGCGGGTCGACGGCCTGGCCGAGCGACCGCTCG 457

Qy      678 GGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCAGCTGCTCCAGGATC 737
      | | ||| || ||||| || || | |||| | ||| ||||
Db      458 ATGCGCTCTCCGGCGGCCAGCGGCAACGCGTGTGGATCGCCATGGTGTGGCGCAGGAAA 517

Qy      738 CTAAGGTCATGCTGTTTGTATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGA 797
      | | | |||| | || ||||| |||| | |||| | | |
Db      518 CCCCCTTGCTGCTGCTCGACGAGCCGACCACCTACCTGGATATCGTCCACCAGATCGAAT 577

Qy      798 TTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTACC 857
      | ||| || || | || || |||| | || | || | || |
Db      578 TGCTCGAACTGCTCGCCGAGCTGAATCGCCAGGGGCGCACCATCGTCGCCGTGCTGCACG 637

Qy      858 AGCCC 862
      | | |
Db      638 ACCTC 642
```

Search completed: February 26, 2004, 09:45:55
Job time : 132.204 secs

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 06:40:42 ; Search time 537.721 Seconds
(without alignments)
15698.623 Million cell updates/sec

Title: US-09-989-981A-5
Perfect score: 2340
Sequence: 1 gtcaggtggagcaggcaggg.....aatattcataaacctatggg 2340

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2353733 seqs, 1803733377 residues

Total number of hits satisfying chosen parameters: 4707466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query				
No.	Score	Match Length	DB	ID		Description

1	2340	100.0	2340	9	US-09-837-992-4	Sequence 4, Appli
2	2340	100.0	2340	10	US-09-989-981A-5	Sequence 5, Appli
3	1395.6	59.6	2258	9	US-09-837-992-2	Sequence 2, Appli
4	1365.4	58.4	1959	10	US-09-989-981A-1	Sequence 1, Appli
5	1174.2	50.2	2512	15	US-10-104-047-825	Sequence 825, App
6	472	20.2	472	9	US-09-837-992-19	Sequence 19, Appl
7	249	10.6	249	9	US-09-837-992-7	Sequence 7, Appli
8	214	9.1	214	9	US-09-837-992-14	Sequence 14, Appl
9	206	8.8	206	9	US-09-837-992-15	Sequence 15, Appl
10	203.6	8.7	2669	10	US-09-989-981A-7	Sequence 7, Appli
11	199.2	8.5	2019	10	US-09-989-981A-3	Sequence 3, Appli
12	186	7.9	186	9	US-09-837-992-17	Sequence 17, Appl
13	156.8	6.7	2585	12	US-10-425-114-32175	Sequence 32175, A
14	140	6.0	140	9	US-09-837-992-12	Sequence 12, Appl
15	139.6	6.0	759	15	US-10-027-632-152155	Sequence 152155,
16	139.6	6.0	759	15	US-10-027-632-152156	Sequence 152156,
17	139.6	6.0	759	15	US-10-027-632-152157	Sequence 152157,
18	139	5.9	139	9	US-09-837-992-16	Sequence 16, Appl
19	137	5.9	137	9	US-09-837-992-9	Sequence 9, Appli
20	135.4	5.8	472	10	US-09-918-995-30637	Sequence 30637, A
21	130	5.6	130	9	US-09-837-992-13	Sequence 13, Appl
22	129	5.5	129	9	US-09-837-992-11	Sequence 11, Appl
c 23	122.2	5.2	371	15	US-10-125-968-1321	Sequence 1321, Ap
24	122	5.2	122	9	US-09-837-992-8	Sequence 8, Appli
25	114.2	4.9	2133	15	US-10-210-130-13	Sequence 13, Appl
26	114.2	4.9	2930	9	US-09-954-531-591	Sequence 591, App
27	114.2	4.9	2930	14	US-10-171-581-276	Sequence 276, App
28	114.2	4.9	2930	16	US-10-429-160-9	Sequence 9, Appli
29	114.2	4.9	3201	13	US-10-072-621-5	Sequence 5, Appli
30	113	4.8	113	9	US-09-837-992-18	Sequence 18, Appl
31	110.8	4.7	2687	13	US-10-154-452-3	Sequence 3, Appli
32	110.8	4.7	2687	14	US-10-090-455-12	Sequence 12, Appl
33	107.6	4.6	2687	13	US-10-154-452-7	Sequence 7, Appli
34	107.2	4.6	1941	14	US-10-090-455-3	Sequence 3, Appli
35	107.2	4.6	3455	13	US-10-072-621-4	Sequence 4, Appli
36	107.2	4.6	3455	14	US-10-090-455-1	Sequence 1, Appli
37	107	4.6	1941	14	US-10-079-087-3	Sequence 3, Appli
38	107	4.6	3526	14	US-10-079-087-1	Sequence 1, Appli
39	103	4.4	103	9	US-09-837-992-10	Sequence 10, Appl
40	101.6	4.3	6043	10	US-09-989-981A-9	Sequence 9, Appli
41	100.4	4.3	3463	12	US-10-425-114-13286	Sequence 13286, A
42	100.4	4.3	3528	12	US-10-424-599-99236	Sequence 99236, A
43	98	4.2	2027	15	US-10-405-806-1	Sequence 1, Appli
44	98	4.2	2053	15	US-10-405-806-12	Sequence 12, Appl
45	98	4.2	2247	9	US-09-866-866A-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
 US-09-837-992-4
 ; Sequence 4, Application US/09837992
 ; Patent No. US20020081687A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tian, Hui

```

; APPLICANT:  Schultz, Joshua
; APPLICANT:  Shan, Bei
; APPLICANT:  Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG):  Compositions
; TITLE OF INVENTION:  and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human sitosterolemia gene (SSG)
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: protein
US-09-837-992-4

```

```

Query Match          100.0%;  Score 2340;  DB 9;  Length 2340;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 2340;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1  GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
      |||
Db      1  GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGGCTCCCCAACTGAAGCCACTCTGGGGA 60

Qy     61  GGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
      |||
Db     61  GGGTCCGGCCACCAGAAAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120

Qy    121  ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
      |||
Db    121  ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180

Qy    181  GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240
      |||
Db    181  GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240

Qy    241  CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC 300
      |||
Db    241  CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC 300

Qy    301  CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT 360
      |||
Db    301  CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT 360

Qy    361  AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG 420
      |||
Db    361  AGGAAGCTCAGGCTCCGGGAAAACACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG 420

```

Qy	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Db	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCG	540
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Qy	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	661	TGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTGACAAAATTGCCAT	900
Db	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTCAGCTCTTGACAAAATTGCCAT	900
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Qy	961	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1020
Db	961	TGACTGCGGTTACCCTTGTCTGAACATTCAAACCCTTTTGACTTCTATATGGACCTGAC	1020
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAAACCAAAGATTCTCCTGGAGTTTT	1200
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260

Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Db	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCT	1320
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Db	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Qy	1501	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1501	GGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGGATATTTTCTGCTGC	1620
Db	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGGATATTTTCTGCTGC	1620
Qy	1621	TCTCTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1680
Db	1621	TCTCTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1680
Qy	1681	AAATCCAAATATAGTCAACAGTGAGTGGCTCTGCTGTCCATTGCGGGGTGCTTGTGG	1740
Db	1681	AAATCCAAATATAGTCAACAGTGAGTGGCTCTGCTGTCCATTGCGGGGTGCTTGTGG	1740
Qy	1741	ATCTGGATTCCCTCAGAAACATACAAGAAATGCCATTCTTTTAAAATCATCAGTTATTT	1800
Db	1741	ATCTGGATTCCCTCAGAAACATACAAGAAATGCCATTCTTTTAAAATCATCAGTTATTT	1800
Qy	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Qy	1861	CACTTGTGGCAGCTCAAATGTTTCTGTGACAATAATCCAATGTGTGCCTTCACTCAAGG	1920
Db	1861	CACTTGTGGCAGCTCAAATGTTTCTGTGACAATAATCCAATGTGTGCCTTCACTCAAGG	1920
Qy	1921	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACAAATGAACCTTCT	1980
Db	1921	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACAAATGAACCTTCT	1980
Qy	1981	GATTTTGTATTCAATTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAG	2040
Db	1981	GATTTTGTATTCAATTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAG	2040
Qy	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Db	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Qy	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTCTTGACAG	2160

```

      |||
Db      2101 ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG 2160
Qy      2161 GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
      |||
Db      2161 GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC 2220
Qy      2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
      |||
Db      2221 TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT 2280
Qy      2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
      |||
Db      2281 TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340

```

RESULT 2

US-09-989-981A-5

```

; Sequence 5, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human ABCG5 (hABCG5)

```

US-09-989-981A-5

```

Query Match      100.0%; Score 2340; DB 10; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
      |||
Db      1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
Qy      61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
      |||

```


Db	61	GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC	120
Qy	121	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Db	121	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT	180
Qy	181	GGAGGGGGCTCCTGCCACCGCCCCGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Db	181	GGAGGGGGCTCCTGCCACCGCCCCGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA	240
Qy	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Db	241	CAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGAC	300
Qy	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Db	301	CAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCT	360
Qy	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Db	361	AGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCG	420
Qy	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Db	421	CGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGTT	480
Qy	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCCTGCG	540
Db	481	CCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCCTGCG	540
Qy	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Db	541	CGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCA	600
Qy	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Db	601	GAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGAT	660
Qy	661	TGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Db	661	TGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGC	720
Qy	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Db	721	CCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTG	780
Qy	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Db	781	CATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGT	840
Qy	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Db	841	GGTTCTCACCATTACACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAAATTGCCAT	900
Qy	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960
Db	901	CCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTTCAA	960

Qy	961	TGACTGCGGTTACCCCTTGTCCTGAACATTCAAACCCCTTTTGACTTCTATATGGACCTGAC	1020
Db	961	TGACTGCGGTTACCCCTTGTCCTGAACATTCAAACCCCTTTTGACTTCTATATGGACCTGAC	1020
Qy	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Db	1021	GTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGATGAT	1080
Qy	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Db	1081	AGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATATTGAAAGAAT	1140
Qy	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Db	1141	GAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTT	1200
Qy	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Db	1201	CTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGC	1260
Qy	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCCTCCTTTCTTCGTTCT	1320
Db	1261	AGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCCTCCTTTCTTCGTTCT	1320
Qy	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Db	1321	GCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCA	1380
Qy	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Db	1381	GTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCT	1440
Qy	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Db	1441	GCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCT	1500
Qy	1501	GGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Db	1501	GGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAGTGT	1560
Qy	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGTTGATATTTTCTGCTGC	1620
Db	1561	GTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCATTGTTGATATTTTCTGCTGC	1620
Qy	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1680
Db	1621	TCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCA	1680
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG	1740
Db	1681	AAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGG	1740
Qy	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTT	1800
Db	1741	ATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTT	1800

Qy	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Db	1801	TACATTCCAAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTT	1860
Qy	1861	CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG	1920
Db	1861	CACTTGTGGCAGCTCAAATGTTTCTGTGACAACATAATCCAATGTGTGCCTTCACTCAAGG	1920
Qy	1921	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACAAATGAACTTTCT	1980
Db	1921	AATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACAAATGAACTTTCT	1980
Qy	1981	GATTTTGTATTCAATTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAG	2040
Db	1981	GATTTTGTATTCAATTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAG	2040
Qy	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Db	2041	GGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCG	2100
Qy	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160
Db	2101	ACTGTGCATGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAG	2160
Qy	2161	GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC	2220
Db	2161	GACATCTCAAGTCTTTTAACCATTAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCC	2220
Qy	2221	TTGAATGCAATGGAAGTGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT	2280
Db	2221	TTGAATGCAATGGAAGTGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGT	2280
Qy	2281	TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2340
Db	2281	TATTTGGAAATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG	2340

RESULT 3

US-09-837-992-2

; Sequence 2, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 2
; LENGTH: 2258
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)
; NAME/KEY: CDS
; LOCATION: (47)..(2005)
; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: protein
US-09-837-992-2
```

```
Query Match          59.6%; Score 1395.6; DB 9; Length 2258;
Best Local Similarity 80.7%; Pred. No. 0;
Matches 1642; Conservative 0; Mismatches 389; Indels 3; Gaps 1;
```

```
Qy      61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
      ||| | ||||| ||||| | ||||| ||| | ||||| ||| |
Db      1 GGGACAGGCCACTAGAAAATTCATTGCAATTTGCTTCCTGCTAGCCATGGGTGAGCTGCC 60

Qy     121 ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
      | | ||| | | | | | | | | | | | | | | | | | |
Db      61 CTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCT 120

Qy     181 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTGGGCATCCTCCATGCCTC 237
      |||| | | | | | | | | | | | | | | | | | |
Db     121 GGAGCAAGGTTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTGTC 180

Qy     238 CTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTG 297
      ||||| ||||| ||| | | | | | | | | | | | | | |
Db     181 CTACAGCGTCAGCAACCGTGTCTGGGCCTTGGTGGAAACATCAAATCATGCCAGCAGAAGTG 240

Qy     298 GACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCAT 357
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db     241 GGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCAT 300

Qy     358 CCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGG 417
      | |||| | ||||| ||||| ||||| ||||| ||||| ||||| ||
Db     301 CTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCG 360

Qy     418 GCGCGCGGGGACCTTCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCA 477
      |||| | ||||| | ||||| ||||| ||||| ||| | | ||||| ||| ||
Db     361 GCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCA 420

Qy     478 GTTCCAGGACTGCTTCTCCTACGTCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCCT 537
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db     421 GTTCCAAGACTGCTTCTCCTACGTCTGCAGAGCGACGTTTTTCTGAGCAGCCTCACTGT 480

Qy     538 GCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTT 597
      ||||| ||||| ||| ||||| ||| ||||| || ||||| || | | | |
Db     481 GCGCGAGACGTTGCGATACAGCGATGCTGGCCCTCTGCCGAGCTCCGCGGACTTCTA 540

Qy     598 CCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACT 657
      | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||| |
Db     541 CAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAAT 600

Qy     658 GATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGCGGGTCTCCATCGC 717
```

Db	601	GATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGC	660
Qy	718	AGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGA	777
Db	661	AGCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGACTGGA	720
Qy	778	CTGCATGACTGCTAATCAGATTGTCTGCTCCTGGTGGAACTGGCTCGCAGGAACCGAAT	837
Db	721	CTGCATGACTGCAAAATCAAATTGTCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAAT	780
Qy	838	TGTGGTTCTCACCATTACCCAGCCCCGTTCTGAGCTTTTTCTAGCTCTTTGACAAAATTGC	897
Db	781	TGTGATTGTCAACCATCCACCAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGC	840
Qy	898	CATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTGATTTCTT	957
Db	841	CATCCTGACTTACGGAGAGTTGGTGTCTGTGGCACCCAGAGGAGATGCTTGGCTTCTT	900
Qy	958	CAATGACTGCGGTTACCCCTTGTCTGAACATTCAAACCCCTTTTGACTTCTATATGGACCT	1017
Db	901	CAATAACTGTGGTTACCCCTGTCTGAACATTCCAATCCCTTTGATTTTTACATGGACTT	960
Qy	1018	GACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAGTCCAGAT	1077
Db	961	GACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTACAAGCGAGTACAGAT	1020
Qy	1078	GATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAAACTTTGAAGAATATTGAAAG	1137
Db	1021	GCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTGGAGAACATTGAAAG	1080
Qy	1138	AATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTCCTGGAGT	1197
Db	1081	AGCACGATACCTGAAAACCTTACCCATGGTTCCTTTCAAACAAAAGATCCTCCTGGGAT	1140
Qy	1198	TTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTGGTGAGAAATAAGCT	1257
Db	1141	GTTCGGCAAGCTTGGTGTCTGCTGAGGCGAGTAACAAGAACTTAATGAGGAATAAGCA	1200
Qy	1258	GGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTCTTCGT	1317
Db	1201	GGCAGTGATTATGCGTCTCGTTCAGAATCTGATCATGGGCCTCTTCCTCATTTTCTACCT	1260
Qy	1318	TCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTA	1377
Db	1261	TCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGCGTGGGGCTGCTCTA	1320
Qy	1378	CCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGT	1437
Db	1321	TCAGCTTGTGGGTGCCACCCCATACACCGGCATGCTCAATGCTGTGAATCTGTTTCCCAT	1380
Qy	1438	GCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGAT	1497
Db	1381	GCTGAGAGCCGTACGCGACCAGGAGAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCT	1440
Qy	1498	GCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCATGATTTTCAGCAG	1557

Db 1441 GCTCGCCTACGTGCTACACGTCTCCCTTCAGCGTCATCGCCACGGTCATTTTCAGCAG 1500
 Qy 1558 TGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATTTGGATATTTTCTGC 1617
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1501 TGTGTGTTATTGGACTCTGGGCTTGATCCTGAAGTTGCCAGATTTGGATATTTCTCTGC 1560
 Qy 1618 TGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGT 1677
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 Db 1561 TGCTCTTTTGGCCCCCTCACTTAATTGGAGAATTTCTAACACTTGTGCTGCTTGGTATAGT 1620
 Qy 1678 CAAAATCCAAATATAGTCAACAGTGTAAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGT 1737
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1621 CAAAACCCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATCTCTGGGCTGCTTAT 1680
 Qy 1738 TGGATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCATCAGTTA 1797
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1681 TGGATCTGGATTTATCAGAAACATACAAGAAATGCCCATTCCTTTTAAATCCTGGGTTA 1740
 Qy 1798 TTTTACATTCCAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAA 1857
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1741 TTTTACATTCCAAAATACTGTTGTGAGATTCTCGTGGTCAATGAGTTTACGGCCGTGAA 1800
 Qy 1858 TTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAATAATCCAATGTGTGCCTTCACTCA 1917
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1801 CTTCACCTTGTGGTGGATCCAACACCTCTATGCTAAATCACCCGATGTGCGCCATCACCCA 1860
 Qy 1918 AGGAATTCAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACAATGAACCTT 1977
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1861 AGGGGTCCAGTTCATCGAGAAAACCTGCCAGGTGCTACATCCAGATTCACGGCAAACCTT 1920
 Qy 1978 TCTGATTTTGTATTCAATTTATTCCAGCTCTTGTGTCATCCTAGGAATAGTTGTTTTCAAAT 2037
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1921 CCTCATCTTATATGGGTTTATCCCAGCTCTGGTGCATCCTAGGAATAGTGATTTTTAAAGT 1980
 Qy 2038 AAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTG 2091
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1981 CAGGGACTACCTGATTAGCAGATAGTTAAGATGACAGGCAGGAAAGGGTTAATG 2034

RESULT 4

US-09-989-981A-1

; Sequence 1, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

; FILE REFERENCE: 018781-007320US

; CURRENT APPLICATION NUMBER: US/09/989,981A

; CURRENT FILING DATE: 2002-07-23

; PRIOR APPLICATION NUMBER: US 60/252,235

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/253,645

```

; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1959)
; OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-1

```

```

Query Match          58.4%; Score 1365.4; DB 10; Length 1959;
Best Local Similarity 81.4%; Pred. No. 0;
Matches 1595; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

```

```

Qy      107 ATGGGTGACCTCTCATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGC 166
          ||||| || | | | || | | || | | | | |||||
Db      1 ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGG 60

Qy      167 TCCCAGAGCTCCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCT---CACAGCCTGGGC 223
          || | ||||| ||||| | | | | || | | || | | || ||
Db      61 TCTCTGAGCTCCCTGGAGCAAGGTTCCGCTCACGGGCACAGAGGCTCGGCACAGCTTAGGT 120

Qy      224 ATCCTCCATGCCTCCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCT 283
          |||| |||| | ||||| ||||| |||| | || | |||| | |||| |||
Db      121 GTCCTGCATGTGTCTACAGCGTCAGCAACCGTGTCTGGGCCTTGGTGGAAACATCAAATCA 180

Qy      284 TGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGG 343
          |||| |||| |||| | |||| | ||||| ||||| ||||| || |||| ||
Db      181 TGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGC 240

Qy      344 CAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATG 403
          |||| | ||||| |||| | ||||| |||| | ||||| ||||| ||||| |||||
Db      241 CAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGACGCCATC 300

Qy      404 TCCGGGAGGCTGGGGCGCGGGGGACCTTCTGGGGGAGGTGTATGTGAACGGCCGGGCG 463
          ||||| |||| | |||| | |||| | |||| | |||| | |||| | || || |
Db      301 TCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAG 360

Qy      464 CTGCGCCGGGAGCAGTTCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTG 523
          |||| | |||| | |||| | |||| | |||| | |||| | |||| | || || |
Db      361 CTGCGCAGGGACCAGTTCGAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTG 420

Qy      524 AGCAGCCTCACCGTGCGGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGC 583
          ||||| |||| | |||| | |||| | |||| | |||| | |||| | || || |
Db      421 AGCAGCCTCACTGTGCGGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGAGC 480

Qy      584 AATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCAT 643
          | | | | | | | ||||| |||| | |||| | |||| | |||| | || || |
Db      481 TCCGCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCAC 540

Qy      644 GTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGCGC 703
          |||| | |||| | |||| | |||| | |||| | |||| | |||| | || || |
Db      541 GTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGC 600

```

Qy	704	CGGGTCTCCATCGCAGCCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCA	763
Db	601	CGAGTTTCCATCGCAGCCCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCA	660
Qy	764	ACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCT	823
Db	661	ACCACAGGACTGGACTGCATGACTGCAAATCAAATTGTCTTCTCTTGGCTGAGCTGGCT	720
Qy	824	CGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTCTGAGCTTTTTTCAGCTC	883
Db	721	CGCAGGGACCGAATTGTGATTGTCAACATCCACCAGCCTCGCTCTGAGCTCTTCCAACAC	780
Qy	884	TTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAA	943
Db	781	TTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTTCTGTGGCACCCAGAGGAG	840
Qy	944	ATGCTTGATTCTTCAATGACTGCGGTTACCCCTGTCTCTGAACATTCAAACCTTTTGAC	1003
Db	841	ATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCTCTGAACATTCCAATCCCTTTGAT	900
Qy	1004	TTCTATATGGACCTGACGTGAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCC	1063
Db	901	TTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTAC	960
Qy	1064	AAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTG	1123
Db	961	AAGCGAGTACAGATGCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTG	1020
Qy	1124	AAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAA	1183
Db	1021	GAGAACATTGAAAGAGCACGATACCTGAAAACCTTACCCATGGTTCCTTTCAAACAAAA	1080
Qy	1184	GATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGAGGAGAGTGACAAGAACTTG	1243
Db	1081	GATCCTCCTGGGATGTTTCGGCAAGCTTGGTGCTCTGCTGAGGCGAGTAACAAGAACTTA	1140
Qy	1244	GTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTT	1303
Db	1141	ATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTTTCAGAATCTGATCATGGGCTCTTC	1200
Qy	1304	CTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGC	1363
Db	1201	CTCATTTTCTACCTTCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGC	1260
Qy	1364	GTAGGTCTCCTTTTACCAGTTTGTGGGCGCCACCCGTACACAGGCATGCTGAACGCTGTG	1423
Db	1261	GTGGGGCTGCTCTATCAGCTTGTGGGTGCCACCCCATACACCGGCATGCTCAATGCTGTG	1320
Qy	1424	AATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAG	1483
Db	1321	AATCTGTTTCCCATGCTGAGAGCCGTCAGCGACCAGGAGAGTCAGGATGGCCTGTATCAT	1380
Qy	1484	AAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCTTCAGCGTTGTTGCCACC	1543
Db	1381	AAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCTCCCTTCAGCGTCATCGCCACG	1440
Qy	1544	ATGATTTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCGATT	1603

Db	1441	GTCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTTGTATCCTGAAGTTGCCAGATTT	1500
Qy	1604	GGATATTTTTCTGCTGCTCTCTTGGCCCCCACTTAATTGGTGAATTTCTAACTCTTGTG	1663
Db	1501	GGATATTTCTCTGCTGCTCTTTTGGCCCCTCACTTAATTGGAGAATTTCTAACTTGTG	1560
Qy	1664	CTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGTGTAGTGGCTCTGCTGTCCATT	1723
Db	1561	CTGCTTGGTATAGTCCAAAACCCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATC	1620
Qy	1724	GCGGGGGTGCTTGTGGATCTGGATTCTCAGAAACATACAAGAAATGCCCATTCCTTTT	1783
Db	1621	TCTGGGCTGCTTATTGGATCTGGATTATCAGAAACATACAAGAAATGCCCATTCCTTTA	1680
Qy	1784	AAAATCATCAGTTATTTTACATTCCAAAATATTGCAGTGAGATTCTTGTAGTCAATGAG	1843
Db	1681	AAAATCCTGGGTATTTTACATTCCAAAATACTGTTGTGAGATTCTCGTGGTCAATGAG	1740
Qy	1844	TTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAATG	1903
Db	1741	TTTTACGGCCTGAACTTCACTTGTGGTGGATCCAACACCTCTATGCTAAATCACCCGATG	1800
Qy	1904	TGTGCCTTCACTCAAGGAATTCATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGA	1963
Db	1801	TGCGCCATCACCCAAGGGGTCCAGTTCATCGAGAAAACCTGCCAGGTGCTACATCCAGA	1860
Qy	1964	TTCACAATGAACTTTCTGATTTTGTATTCAATTTATTCCAGCTCTTGTATCCTAGGAATA	2023
Db	1861	TTCACGGCAAACCTCCTCATCTTATATGGGTTTATCCCAGCTCTGGTCATCCTAGGAATA	1920
Qy	2024	GTTGTTTTTCAAATAAGGGATCATCTCATTAGCAGGTAG	2062
Db	1921	GTGATTTTAAAGTCAGGGACTACCTGATTAGCAGATAG	1959

RESULT 5

US-10-104-047-825

; Sequence 825, Application US/10104047

; Publication No. US20030236392A1

; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA

; FILE REFERENCE: H1-A0105

; CURRENT APPLICATION NUMBER: US/10/104,047

; CURRENT FILING DATE: 2002-03-25

; PRIOR APPLICATION NUMBER:

; PRIOR FILING DATE:

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 825

; LENGTH: 2512

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-104-047-825

Query Match

50.2%; Score 1174.2; DB 15; Length 2512;

Best Local Similarity 71.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 603; Indels 103; Gaps 9;

```
Qy      1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 140

Qy     61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 200

Qy    121 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 260

Qy    181 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 320

Qy    241 CAGCGTCAGCCACCGC----GTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGT 296
      ||||||||  |  ||          ||  |  |  |  |  |||  ||  |
Db    321 CAGCGTCAGGTAAGGCAGAGCCCTTGCTGCTGCTGCTCCCCAGGAGTGCGGGGCCCGGC 380

Qy    297 GGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCA 356
      |  |  |  |  |  |  |  |  |||  |  |  |||  ||  |  |
Db    381 GCTCACCCCTCTGCTGCCTTTCTTCACTCTTTAAGTGCCAGTCTGGGCACTTCGGGCTCC 440

Qy    357 TCCTAGGAAGCTCAGGCTCC-----GGGAAAACCACGCTGCTGGACGCCATGTCC 406
      ||  |  |  ||  |  ||||  ||  |  ||||  ||||
Db    441 CTCTTTAGTGGATCGGGTGGAGAGAGGAGAGGGAGAAGGGCTGTGCTGGGAAACATGGAG 500

Qy    407 GGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTG 466
      |  |  |  |  |  |  |  ||||  ||  |  |  |  |  |||
Db    501 CGACAGTGAATGGCCCCCTCCCCCTGCCAGGGAAGGGCCTGGGCATAAACAAAGTGGCAG 560

Qy    467 CGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGAGC 526
      |  |  |  ||  |  |  ||||  |||  ||  |  |  |
Db    561 CAGTGCCCTGCCAACCCAGTGTCTACGGCCTGCCCTCTGTGGATGGGAATGGGGTACTG 620

Qy    527 AGCCT-----CACCGTGCGCAGACGCTGCACTACACCGCGCTGCTGGCCAT 573
      |  |  |  ||||  |  ||||  ||  |  ||  ||||
Db    621 CGAATGCAAGGAGTCTTGAAACCTGGTGAAAGAATGCAGGGACAGCCACCTCGCAGCCAA 680

Qy    574 CCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCAGAGCTGAG 633
      ||  ||  ||  ||  |  ||||  ||  |  ||  |||||
Db    681 ACGGACAGGACATTTCAGAGCAACTCCAGCACAGGCCCCCTCCCTACGTGGCAGACAGCCT 740

Qy    634 TCTGAGCCATGTGGCAG-----ACCGACTGATTGGCAACTACAGCT 674
      ||  ||  ||  ||  |  ||||  |  ||||  |||
Db    741 CAGTCGCTATCTGCCAGGTTCTACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCC 800

Qy    675 TGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCAT-----CG 716
      ||  ||  |  ||||  |  |  |  ||  ||
Db    801 TGTCCGAGACTACTGGGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTGGGTGTGGC 860

Qy    717 CAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGG 776
      |  |  ||  ||  ||  |  ||  ||  |  ||  ||  |  |
```

Db 861 CCCTTCCAGGGCCCCAAGCTGCCTTTGCCTTCCTGGGGTTTCCTTTAAAGCCACCGCGTG 920
 Qy 777 ACTGCATGACTGCTAATCAGATTGTCGTCTCTGGTGGAACTGGCTCGCAGGAACCGAA 836
 | | | | | | | | | | | | | | | | | | | | | |
 Db 921 AGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGAT 980
 Qy 837 TTGTGGTTCTCACCATTACCAGCCCCGTTCTGAGCTTTTTCAG-----CTCTTTGAC 889
 | | | | | | | | | | | | | | | | | | | | | |
 Db 981 GTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGG 1040
 Qy 890 AAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTT 949
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1041 AAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGG 1100
 Qy 950 GATTTCTTCAATGACTGCGGTTACCCCTGTCTGAACATTCAAACCCCTTTGACTTCTAT 1009
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1101 GAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGCAGT---TCCAGGACTGCTTCTCC 1157
 Qy 1010 ATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGA 1069
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1158 TACGTCTGCAGAGCGACACCCTGCTGAGCAGCCTCACCCTGCGCGAGACGCTGCACTAC 1217
 Qy 1070 GTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCAAAAACCTTTGAAGAAT 1129
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1218 ACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCC 1277
 Qy 1130 AT-----TGAAAGAATGAAACACCTGAAAACGTTACCAA 1163
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1278 GTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTG 1337
 Qy 1164 TGGTTCCTTTCAAACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCCTGA 1223
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1338 GGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGAT 1397
 Qy 1224 ---GGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTC 1280
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1398 CCTAGAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTC 1457
 Qy 1281 AGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGC 1340
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1458 AGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGC 1517
 Qy 1341 TAAAGGGTGCTATCCAGGACCGCTAGGTCTCCTTTTACCAGTTTGTGGGCGCCACCCCGT 1400
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1518 TAAAGGGTGCTATCCAGGACCGCTAGGTCTCCTTTTACCAGTTTGTGGGCGCCACCCCGT 1577
 Qy 1401 ACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGG 1460
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1578 ACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGG 1637
 Qy 1461 AGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCC 1520
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1638 AGAGTCAGGACGGCCTCTACCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCC 1697
 Qy 1521 TCCCTTCAGCGTTGTTGCCACCATGATTTTTCAGCAGTGTGTGCTACTGGACGCTGGGCT 1580
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1698 TCCCTTCAGCGTTGTTGCCACCATGATTTTTCAGCAGTGTGTGCTACTGGACGCTGGGCT 1757

Qy	1581	TACATCCTGAGGTTGCCCCGATTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCTTAA	1640
Db	1758	TACATCCTGAGGTTGCCCCGATTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCTTAA	1817
Qy	1641	TTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACA	1700
Db	1818	TTGGTGAATTTCTAACTCTTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACA	1877
Qy	1701	GTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGGATCTGGATTCCTCAGAAACA	1760
Db	1878	GTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGGATCTGGATTCCTCAGAAACA	1937
Qy	1761	TACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAATATTGCA	1820
Db	1938	TACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAATATTGCA	1997
Qy	1821	GTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATG	1880
Db	1998	GTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCTCAAATG	2057
Qy	1881	TTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAA	1940
Db	2058	TTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATTCATTGAGAAAA	2117
Qy	1941	CCTGCCCAGGTGCAACATCTAGATTACAAATGAACCTTTCTGATTTTGTATTCAATTTATTC	2000
Db	2118	CCTGCCCAGGTGCAACATCTAGATTACAAATGAACCTTTCTGATTTTGTATTCAATTTATTC	2177
Qy	2001	CAGCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGT	2060
Db	2178	CAGCTCTTGTATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATCTCATTAGCAGGT	2237
Qy	2061	AGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGA	2120
Db	2238	AGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGA	2297
Qy	2121	ACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTCAAGTCTTTTAAC	2180
Db	2298	ACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTCAAGTCTTTTAAC	2357
Qy	2181	CATTAAGACTCCATTTGTGCCTCTTGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGT	2240
Db	2358	CATTAAGACTCCATTTGTGCCTCTTGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGT	2417
Qy	2241	TTATAGTCCCTTGCTCTTACAACTTGCAGGGACATGTGGTTATTTGGAAATTGTGACTGA	2300
Db	2418	TTATAGTCCCTTGCTCTTACAACTTGCAGGGACATGTGGTTATTTGGAAATTGTGACTGA	2477
Qy	2301	GCGGACCCAAGAATGTAAATAATATTCATAAACCT	2335
Db	2478	GCGGACCCAAGAATGTAAATAATATTCATAAACCT	2512

RESULT 6

US-09-837-992-19

; Sequence 19, Application US/09837992

```

; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 13 of hSSG
US-09-837-992-19

```

```

Query Match          20.2%; Score 472; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.2e-131;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1869 GCAGCTCAAATGTTTCTGTGACAACCTAATCCAATGTGTGCCTTCACTCAAGGAATTCAAT 1928
          |||
Db      1 GCAGCTCAAATGTTTCTGTGACAACCTAATCCAATGTGTGCCTTCACTCAAGGAATTCAAT 60

Qy      1929 TCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTACAAATGAACTTTCTGATTTTGT 1988
          |||
Db      61 TCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTACAAATGAACTTTCTGATTTTGT 120

Qy      1989 ATTCATTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATC 2048
          |||
Db      121 ATTCATTTATTCCAGCTCTTGTGCATCCTAGGAATAGTTGTTTTCAAATAAGGGATCATC 180

Qy      2049 TCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCA 2108
          |||
Db      181 TCATTAGCAGGTAGTGAAAGCCATGGCTGGGAAAATGGAAGTGAAGCTGCCGACTGTGCA 240

Qy      2109 TGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTC 2168
          |||
Db      241 TGACTGCTCTGAACGTCTGAAATGAGAGTGCCATGTATTTCTTTCTTGACAGGACATCTC 300

Qy      2169 AAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGC 2228
          |||
Db      301 AAGTCTTTTAACCATTAAAGACTCCATTTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGC 360

Qy      2229 AATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGTTATTTGGA 2288
          |||
Db      361 AATGGAAGTGGTTTATAGTCCCTTGCTCTTACAACCTGCAGGGACATGTGGTTATTTGGA 420

```

Qy 2289 AATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 2340
|||||
Db 421 AATTGTGACTGAGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGG 472

RESULT 7

US-09-837-992-7

; Sequence 7, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 7

; LENGTH: 249

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: exon 1 of hSSG

US-09-837-992-7

Query Match 10.6%; Score 249; DB 9; Length 249;

Best Local Similarity 100.0%; Pred. No. 2.3e-64;

Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60
|||||
Db 1 GTCAGGTGGAGCAGGCAGGGCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGA 60

Qy 61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120
|||||
Db 61 GGGTCCGGCCACCAGAAAATTTGCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTC 120

Qy 121 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180
|||||
Db 121 ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCT 180

Qy 181 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240
|||||
Db 181 GGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTA 240

Qy 241 CAGCGTCAG 249
|||||
Db 241 CAGCGTCAG 249

RESULT 8

US-09-837-992-14

```
; Sequence 14, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 8 of hSSG
US-09-837-992-14
```

```
Query Match          9.1%; Score 214; DB 9; Length 214;
Best Local Similarity 100.0%; Pred. No. 8e-54;
Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1011 TGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAG 1070
          |||
Db        1 TGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAACCTCCAAGAGAG 60

Qy      1071 TCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATA 1130
          |||
Db        61 TCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTCATAAACTTTGAAGAATA 120

Qy      1131 TTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTC 1190
          |||
Db       121 TTGAAAGAATGAAACACCTGAAAACGTTACCAATGGTTCCTTTCAAACCAAAGATTCTC 180

Qy      1191 CTGGAGTTTCTCTAAACTGGGTGTTCTCCTGAG 1224
          |||
Db       181 CTGGAGTTTCTCTAAACTGGGTGTTCTCCTGAG 214
```

RESULT 9

US-09-837-992-15

```
; Sequence 15, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
```

```
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 9 of hSSG
US-09-837-992-15
```

```
Query Match          8.8%; Score 206; DB 9; Length 206;
Best Local Similarity 100.0%; Pred. No. 2.1e-51;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1225 GAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAA 1284
          |||
Db      1 GAGAGTGACAAGAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGTCTCCTTCAGAA 60

Qy      1285 TCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAA 1344
          |||
Db      61 TCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCGAAGCAATGTGCTAAA 120

Qy      1345 GGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTACAC 1404
          |||
Db      121 GGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTTTGTGGGCGCCACCCCGTACAC 180

Qy      1405 AGGCATGCTGAACGCTGTGAATCTGT 1430
          |||
Db      181 AGGCATGCTGAACGCTGTGAATCTGT 206
```

RESULT 10

US-09-989-981A-7

```
; Sequence 7, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
```


; PRIOR APPLICATION NUMBER: US 60/252,235
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/253,645
 ; PRIOR FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 2669
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (100)..(2121)
 ; OTHER INFORMATION: human ABCG8 (hABCG8)
 US-09-989-981A-7

Query Match 8.7%; Score 203.6; DB 10; Length 2669;
 Best Local Similarity 54.4%; Pred. No. 6.8e-50;
 Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

Qy	285	GCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGC	344
Db	335	GCTGCCAGAATTCTTGTGAGCTGGGCATCCAGAACCTAAGCTTCAAAGTGAGAAGTGGGC	394
Qy	345	AGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGT	404
Db	395	AGATGCTGGCCATCATAGGGAGCTCAGGTTGTGGGAGAGCCTCCTTGCTAGATGTGATCA	454
Qy	405	CCGGGAGGCTGGGGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGC	464
Db	455	CTGGCCGAGGTCACGCGCGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCCA	514
Qy	465	TGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACCCTGCTGA	524
Db	515	GCTCGCCTCAGCTGGTGAGGAAGTGTGTGGCCACGTGCGCCAGCACAAACAGCTGCTCC	574
Qy	525	GCAGCCTCACCGTGCGGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCCGGGCA	584
Db	575	CCAACTTGACTGTGCGAGAGACCTTGGCCTTCATTGCCAGATGCGGCTGCCAGAACCT	634
Qy	585	ATCCCGGCTCCTTCC---AGAAGAAGGTGGAGGCCGTATGGCAGAGCTGAGTCTGAGCC	641
Db	635	TCTCCAGGCCAGCGTGACAAAAGGGTGGAGGACGTGATCGCGGAGCTGCGGCTTAGGC	694
Qy	642	ATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACGGGTGAGCGGC	701
Db	695	AGTGCGCTGACACCCGCGTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGTGAAGCGCA	754
Qy	702	GCCGGGTCTCCATCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGC	761
Db	755	GGAGAGTCAGCATTGGGGTGAGCTCCTGTGGAACCCAGGAATCCTTATCTCGACGAAC	814
Qy	762	CAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGG	821
Db	815	CCACCTCTGGGCTCGACAGCTTCACAGCCCACAACCTGGTGAAGACCTGTCCAGGCTGG	874
Qy	822	CTCGCAGGAACCGAATTGTGGTTCTCACCATTACCAGCCCCGTCTGAGCTTTTTCAGC	881

```

      | | | | | | | | | | | | | | | | | | | | | | | |
Db      875 CCAAAGGCAACCGGCTGGTGCTCATCTCCCTCCACCAGCCTCGCTCTGACATCTTCAGGC 934

Qy      882 TCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGG 941
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      935 TGTTTGATCTGGTCCTCCTGATGACGTCTGGCACCCCATCTACTTAGGGGCGGCCAGC 994

Qy      942 AAATGCTTGATTCTTCAATGACTGCGGTTACCCCTTGTCCTGAACATTCAAACCCTTTTG 1001
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      995 ACATGGTCCAGTATTTACAGCCATCGGCTACCCCTGTCCTCGCTACAGCAATCCTGCTG 1054

Qy      1002 ACTTCTATATGGACCTGACGTGAGTACCCAAAGCAAGGAACGGGAAATAGAAACCT 1061
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      1055 ACTTCTATGTGGACCTGACCAGCATTGACAGGCGCAGCAGAGAGCAGGAATTGGCCACCA 1114

Qy      1062 CCAAGAGAGTCCAG 1075
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      1115 GGGAGAAGGCTCAG 1128

```

RESULT 11

US-09-989-981A-3

```

; Sequence 3, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2019)
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-3

```

```

Query Match          8.5%; Score 199.2; DB 10; Length 2019;
Best Local Similarity 54.0%; Pred. No. 1.2e-48;
Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

```

```

Qy      283 TTGCCGGCAGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGACGTGGAGAGCGG 342
      | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 12
US-09-837-992-17

```
; Sequence 17, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 186
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 11 of hSSG
US-09-837-992-17
```

```
Query Match          7.9%; Score 186; DB 9; Length 186;
Best Local Similarity 100.0%; Pred. No. 2.1e-45;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1570 GACGCTGGGCTTACATCCTGAGGTTGCCCATTGATATTTTCTGCTGCTCTCTTGGC 1629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 GACGCTGGGCTTACATCCTGAGGTTGCCCATTGATATTTTCTGCTGCTCTCTTGGC 60

Qy      1630 CCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAA 1689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 CCCCCACTTAATTGGTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAA 120

Qy      1690 TATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATT 1749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            121 TATAGTCAACAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTGGATCTGGATT 180

Qy      1750 CCTCAG 1755
          |||||
Db            181 CCTCAG 186
```

RESULT 13

```
US-10-425-114-32175
; Sequence 32175, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
```

```

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32175
; LENGTH: 2585
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73274A02_FLI
US-10-425-114-32175

```

```

Query Match          6.7%; Score 156.8; DB 12; Length 2585;
Best Local Similarity 51.1%; Pred. No. 9.6e-36;
Matches 424; Conservative 0; Mismatches 397; Indels 9; Gaps 2;

```

```

Qy      177 CCCTGGAGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCT 236
        |||| | ||| | |||| | | ||| | | ||| | |
Db      441 CCCTGTGGCGGGACAGCAAGGCGCTCCCGCCGGGGGCGGCCCGCCGCGCTCATCGGCG 500

Qy      237 CCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGT 296
        | || ||| | | || | | || | | | || | |
Db      501 ACGTGTCCGCCAGGCTCACGTGGAAGGACCTCTGCGTCACCGTGGCTCTGGGCCCCGGCA 560

Qy      297 GGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCA 356
        ||| | | || | || || || || || || || || ||
Db      561 AGACGCAGACCGTGCTGGACGAGCTACCGGGTACGCGGAGCCCGGGTCGCTGACCGCGC 620

Qy      357 TCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGAGGCTGG 416
        || | || || || || || || || || || || || || ||
Db      621 TCATGGGGCCCTCGGGGTCCGGCAAGTCCACCCTGCTCGACGCCCTCGCCGGCCGCTCG 680

Qy      417 GCGCGCGGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCCGGGAGC 476
        | | | || || | || | || || || || || | . | |||
Db      681 CCGCCAACGCCTTCCTCTCCGGCAACGTGCTCCTCAACGGCCGCAAG-----GCCAAGC 734

Qy      477 AGTTCCAGGACTGCTTCTCCTACGTCTGCGAGAGCGACACCCTGCTGAGCAGCCTCACCG 536
        | | | | | || || || || || || || || || || || ||
Db      735 TCTCCTTCGGCGCCGCGGCGTACGTGACGAGGACGACAACCTGATCGGGACGCTGACGG 794

Qy      537 TGCGCGAGACGCTGCACTACACCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCT 596
        |||| |||| | |||| | |||| || | | | | |
Db      795 TGCGCGAGACGATCGGCTACTCGGCGCTGCTGCGGCTGCCGGACAAGATGCCGCGGGAGG 854

Qy      597 TCCAG---AAGAAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACC 653
        | || | |||| || | || | || || || || | || ||
Db      855 ACAAGCGCGCGCTGGTGGAGGGCACCATCGTCGAGATGGGGCTGCAGGACTGCGCCGACA 914

Qy      654 GACTGATTGGCAACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCA 713
        | || |||| || | | || | | | || || || || || ||
Db      915 CCGTCATCGGCAACTGGCACCTCCGCGGGGTGAGCGGCGGCGAGAAGCGCCGCGTCAGCA 974

Qy      714 TCGCAGCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCC 773

```

```

      |||| | |||| ||| | || || | | || |||| |||| ||||
Db      975 TCGCGCTCGAGCTACTCATGCGCCCGCGCCTCCTCTTCCTCGACGAGCCCACCAGCGGCC 1034
Qy      774 TGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACC 833
      | ||| || | |||| | || || |||| | |
Db     1035 TCGACAGCTCGTCTGCGTTCTTCGTGACGCAGACGCTGCGGGGCTGGCGAGGGACGGCA 1094
Qy      834 GAATTGTGGTTCTCACCATTCAACAGCCCCGTTCTGAGCTTTTTTCAGCTCTTTGACAAA 893
      | | ||| || |||| |||| |||| || | || | |||| ||||
Db     1095 GGACGGTGATTGCTTCCATCCACCAGCCCAGCAGCGAGGTGTTTCGAGCTCTTCGACATGC 1154
Qy      894 TTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCAGCCAGCGGAAATGCTTGATT 953
      | | | || | || || | || || || || || || || ||
Db     1155 TCTTCCTGCTATCCGGGGCAAGACCGTCTACTTCGGACAAGCATCGCAAGCATGCGAGT 1214
Qy      954 TCTTCAATGACTGCGGTTACCCCTTGTCCTGAACATTCAAACCCTTTTGAC 1003
      |||| | | |||| |||| || | | || | | ||
Db     1215 TCTTTGCTCAAGCCGGTTTCCCTTGCCCCGGCTCTGCGGAATCCGTCCGAC 1264

```

RESULT 14

US-09-837-992-12

; Sequence 12, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 12

; LENGTH: 140

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: exon 6 of hSSG

US-09-837-992-12

Query Match 6.0%; Score 140; DB 9; Length 140;

Best Local Similarity 100.0%; Pred. No. 1.4e-31;

Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      741 AGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTG 800
      |||||||||
Db      1 AGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACTGCTAATCAGATTG 60
Qy      801 TCGTCCTCCTGGTGGAAGTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTCAACAGC 860

```

```

|||||
Db      61 TCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTGGTTCTCACCATTCACCAGC 120

Qy      861 CCCGTTCTGAGCTTTTTCAG 880
|||||
Db      121 CCCGTTCTGAGCTTTTTCAG 140

```

RESULT 15

US-10-027-632-152155

; Sequence 152155, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 152155

; LENGTH: 759

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-152155

Query Match 6.0%; Score 139.6; DB 15; Length 759;

Best Local Similarity 99.3%; Pred. No. 6.4e-31;

Matches 139; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1431 TTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGC 1490
|||||
Db      42 TTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGC 101

Qy      1491 AGATGATGCTGGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTT 1550
|||||
Db      102 AGATGATGCTGGCCTATGCACTGCACGTCTCCCTTCAGCGTTGTTGCCACCATGATTT 161
Qy      1551 TCAGCAGTGTGTGCTACTGG 1570
|||||
Db      162 TCAGCAGTGTGTGCTACTGG 181

```

Search completed: February 27, 2004, 07:11:39

Job time : 541.721 secs